UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR

Grado en Ingeniería Informática

TRABAJO FIN DE GRADO

DETECCIÓN DE COMUNIDADES EN REDES SOCIALES, ESTUDIO SOBRE LAS ELECCIONES CATALANAS Y ELECCIONES GENERALES DE 2015

Adrián Lorenzo Mateo
Tutor: Pablo Castells Azpilicueta

Julio 2016

DETECCIÓN DE COMUNIDADES EN REDES SOCIALES, ESTUDIO SOBRE LAS ELECCIONES CATALANAS Y ELECCIONES GENERALES DE 2015

AUTOR: Adrián Lorenzo Mateo

TUTOR: Pablo Castells Azpilicueta

Escuela Politécnica Superior
Universidad Autónoma de Madrid
Julio de 2016

Resumen (castellano)

Twitter se ha convertido en un aspecto más de la web social como un reflejo de los temas de actualidad, permitiendo a los usuarios compartir ideas, noticias u opiniones. En materia de comunicación política cada vez más se está utilizando esta plataforma de microblogging como herramienta, convirtiéndose en objeto de análisis de los investigadores para averiguar su validez como indicador del comportamiento u opinión política de la población. Esta contribución se centra en el periodo de campaña en España durante las elecciones de Cataluña de Septiembre de 2015 (27s) y las elecciones generales de Diciembre de 2015 (20d), siguiendo la línea de anteriores investigaciones, contrastándolas y ampliándolas ligeramente. El objetivo de este trabajo es el análisis de la presencia de comunidades latentes a través de las distintas relaciones definidas en Twitter, sus propiedades y particularidades, la forma en la que estos subgrupos interactúan y tienen vínculos, su adecuación como comunidades de simpatizantes políticos, la correlación entre sus características y la estimación del voto y el estudio de la polaridad de opinión en el contexto de la independencia de Cataluña durante las elecciones. Para este estudio se extrajeron de Twitter alrededor de un millón y medio de tweets en el total de ambos escenarios de campaña electoral.

Abstract (English)

Twitter has become an aspect of the social web as a reflection of the current issues, and allowing users to share ideas, news or opinions. This microblogging platform is being increasingly used as a political communication tool, becoming a subject of analysis for researchers, as to determine its validity as an indicator of behavior or political opinion of the population. This contribution focuses on the campaign period for Catalan parliamentary elections on September 2015 (27S) and Spanish general elections on December 2015 (20D), following the line of previous research, contrasting and extending them slightly. The aim of this paper is the analysis of the presence of latent communities through various defined relationships on Twitter, their properties and characteristics, the way that these subgroups interact and how they have connecting links, their fitness as communities of political supporters, the correlation between its characteristics and the vote estimate and the study of the opinion polarity in the context of the independence of Catalonia during the Catalan election. For this study we've extracted from Twitter about one million and a half tweets in total involving both scenarios during the election campaign.

Palabras clave (castellano)

Análisis de redes sociales, Twitter, detección de comunidades, predicción de votos, minería de textos, polaridad de opinión.

Keywords (inglés)

Social network analysis, Twitter, community detection, vote prediction, text mining, opinion polarity.

INDICE DE CONTENIDOS

INDICE DE FIGURAS

INDICE DE TABLAS

1 Introducción

1.1 Motivación

A lo largo de estos últimos años las redes sociales online han comenzado a cobrar un gran protagonismo en las relaciones humanas. Esto ha impulsado que múltiples investigadores hayan empleado estas redes como herramientas para predecir ciertos eventos o acontecimientos, así como para extraer la opinión o el sentir del imaginario colectivo. Sin embargo, este tema no está ausente de debate, ya que algunos puntos de vista advierten de cierta "euforia desmedida" a la hora de afirmar que los datos sociales on-line pueden otorgar la capacidad de predecir ciertos sucesos o tendencias. Y son estas críticas las que ponen en manifiesto que hay que tener en cuenta que las redes sociales son una muestra no siempre representativa de la sociedad. Existen sesgos como la edad, el género, raza o grupo social en estas redes. En general, los jóvenes y las personas de zonas urbanas de clase media-alta están sobre-representados en las redes sociales. Otros fenómenos que producen distorsión son las minorías ruidosas frente a las mayorías silenciosas, la veracidad de las opiniones debido a la presencia de recursos propagandísticos o información engañosa (cuentas automáticas, spammers, etc) (Gayo-Avello, 2012). Una anécdota clara que ejemplifica estos fenómenos es el caso de las elecciones parlamentarias de Alemania del año 2009, en el que según el modelo propuesto por Tumasjan (Tumasjan A. et al, 2010). el Partido Pirata habría ganado 34,8 en lugar de un 2,1 por ciento de los votos (Jungherr et al., 2011).

Sin embargo, a pesar de los sesgos, las redes sociales son valiosas fuentes de datos de opiniones que se pueden utilizar como indicadores de apoyo, a modo de encuestas y sondeos. No se trata de utilizar ingenuamente las opiniones observables en las redes sociales como información fiable y veraz, pero sí pueden ayudar a configurar un "espejo" de la opinión pública reflejado en la interacción entre los usuarios de estas redes. Aunque en Twitter se muestren tendencias y no sondeos, utilizar esta plataforma como herramienta tiene múltiples ventajas: coste cercano a cero, las opiniones se proporcionan de forma voluntaria y gratuita, un gran volumen de datos en tiempo real y de forma prolongada en el tiempo. En relación al aspecto político, hay investigadores que afirman que *"Twitter es una fuente de información que permite segmentar a los usuarios, descubrir cómo los ciudadanos participan en el debate político y como se agrupan por afinidad ideológica"* (Congosto et al., 2013). El análisis de las redes sociales es un campo difícil, es posible que las herramientas se hayan creado primero (las redes sociales) y ahora se deba buscar su aplicación práctica, pero es algo que merece la pena estudiar a pesar de que en muchas ocasiones no se acierte con los resultados.

1.2 Objetivos

El presente documento tiene como propósito general el análisis de la formación de grupos de partidarios políticos dentro de las redes sociales, así como la influencia y tendencia de los mismos en el marco de las elecciones de la comunidad de Cataluña y generales de 2015 a través de la información que circulaba sobre la red social Twitter. Este análisis comprende el estudio de los vínculos entre las comunidades de partidarios, es decir, la posible afinidad o antagonismo entre las mismas, la formación de polaridades de opinión y la valoración del sentimiento político. Asimismo se va a estudiar la posible relación entre las características de las comunidades con la intención de voto. La siguiente lista indica de forma detallada cada uno de los objetivos que se abordarán en el trabajo:

- Construcción de una aplicación recolectora de datos mediante el API Pública de Twitter a través del método Streamming, en torno a búsquedas prefijadas por palabras clave.
- Detección de comunidades de partidarios políticos.
- Análisis de la relación entre las comunidades de partidarios.
- Predicción de la intención de voto a través de las características de las comunidades.
- Análisis de la polaridad de los tweets a través de clasificadores.

1.3 Organización de la memoria

El presente documento se estructura de la siguiente forma:

- **Capítulo 1: Introducción.** Explicación de las motivaciones y objetivos del estudio desarrollado.
- **Capítulo 2: Estado del arte.** Se presentan las diferentes métricas y métodos existentes en el panorama actual para la detección de comunidades en grafos.
- **Capítulo 3: Diseño y Desarrollo.** Se detallan las herramientas utilizadas, la captura del conjunto de tweets sobre los que se trabajará posteriormente, el volumen de datos, el período de extracción, etc. Asimismo, en este apartado se explican los procedimientos utilizados para la obtención de esos conjuntos de datos.
- **Capítulo 4: Resultados.** En este capítulo se presentan las características y propiedades de cada uno de los grafos obtenidos a través de los conjuntos de datos recogidos. También se presenta el resultado de la detección de comunidades en los grafos, la determinación del carácter político de las mismas, los vínculos entre los grupos y su relación con la intención de voto, así como la polaridad de opinión.
- **Capítulo 5: Conclusiones y trabajo futuro.** Por último, en este capítulo, se expone un resumen con las contribuciones del documento así como las futuras líneas de trabajo.
- **Anexo A: Estadísticas de Twitter sobre el conjunto de datos de las Elecciones de Cataluña.**
- **Anexo B: Estadísticas de Twitter sobre el conjunto de datos de las Elecciones Generales.**
- **Anexo C: Grafos con las grupos de partidarios resultantes de la detección de comunidades.**

2 Estado del arte

2.1 Análisis de redes sociales

El estudio de las redes sociales es abordado por diferentes campos del conocimiento como la sociología, la economía o la informática. En estos estudios los nodos de los grafos representan personas o individuos mientras que los enlaces representan las relaciones que llevan a cabo entre sí, pudiendo ser éstas de distinta índole; amistosas, de carácter laboral, etc.

El análisis de las redes sociales se centra en entender, categorizar y cuantificar las estructuras de estas redes, ya que la manera en la que estén configuradas afecta a los individuos y las relaciones que se dan dentro de ellas. Tiene origen en las ciencias sociales y en la aplicación del campo de las matemáticas denominado "teoría de grafos" que proporciona gran cantidad de métodos y métricas abstractas para analizar redes. Por ejemplo, mediante estas métricas se puede conocer dentro de una red social quién es el individuo idóneo para la difusión de la información, quiénes son los formadores de opinión, quiénes son los sujetos más populares o detectar subgrupos de individuos con mayor concentración de relaciones entre sí.

2.2 Twitter

Twitter es una red social de microblogging lanzada en 2006, en la que cada usuario puede publicar mensajes llamados "tweets" de un máximo de 140 caracteres. Esta red social ofrece un enorme flujo de información en tiempo real (gran parte este flujo es de dominio público), sobre diversos temas, noticias, anuncios, política, etc. a través de múltiples formatos, tales como frases, imágenes o videos de corta duración. En cuanto a la estructura de las relaciones en Twitter, los usuarios se siguen los unos a los otros (relación follow), pero no necesariamente de una forma bidireccional, por lo que constituye una red social dirigida.

Además, los usuarios en Twitter pueden interactuar entre sí mediante varias formas de comunicación: la mención (quote) consiste en referenciar a un usuario escribiendo en un tweet el símbolo @ seguido del alias del usuario para que así se dé por aludido; el retweet (RT), que consiste en reenviar un determinado tweet a todos los seguidores, éste puede ir acompañado de un comentario, por lo que se recomienda no usar los 140 caracteres cuando se envía un twee; la respuesta (reply) que es la respuesta pública a una mención o retweet; y los llamados hashtags, que son palabras clave con el símbolo # delante, las cuales permiten agrupar, ver y buscar tweets que contengan esa misma palabra clave. Twitter también

permite la geolocalización de los tweets a través de las coordenadas geográficas del lugar donde se realizaron.

Año tras año Twitter se utiliza cada vez más como herramienta política, en campaña electoral los políticos aumentan su presencia en las redes sociales comunicándose con el electorado a la vez que intentan ganar votos aumentando su notoriedad. Como resultado, las redes sociales, y en concreto Twitter, ofrecen un reflejo del panorama político real.

2.3 Redes libres de escala

Las redes complejas pueden estructurarse de diversas formas, siendo la topología libre de escala una de las más extendidas. En una red libre de escala, algunos nodos están altamente conectados, es decir, poseen un gran valor de grado (número de enlaces a otros nodos), aunque el número de conexiones de casi todos los nodos es bastante bajo. En estas redes los nuevos nodos que se añadan tienen mayor probabilidad de enlazarse de forma con los nodos que concentren mayor número de enlaces. Mientras que en una red aleatoria la distribución del grado de los nodos sigue una campana de Gauss (la gran mayoría de los nodos tienen el mismo grado), la distribución de los nodos de las redes libres de escala sigue una ley de potencias. Una distribución de grado siguiendo la ley de potencias implica que pocos nodos concentran la mayor parte de los enlaces, mientras que la mayoría de los nodos sólo tienen unos pocos. Tienen esa estructura por ejemplo: las redes sociales, redes metabólicas y de proteínas, genéticas, etc. Las redes libres de escala y sus propiedades fueron descubiertas por Lászlo Barabási (Barabási & Albert, 1999) junto con sus colaboradores al realizar un mapa de la Web. Estas redes son bastante robustas ante la eliminación aleatoria de nodos. Se puede eliminar el 80% de los nodos y que la red continúe funcionando con su estructura. Por otro lado estas redes son muy débiles si se eliminan nodos concretos con un alto grado, lo que provocaría una fragmentación en la estructura de la red, convirtiéndose en un conjunto de grafos aislados. Estos fenómenos fueron estudiados por Cohen y otros a través de la llamada teoría de la percolación (Cohen et al., 2000). La Figura 1 ofrece una explicación gráfica del proceso de fragmentación.

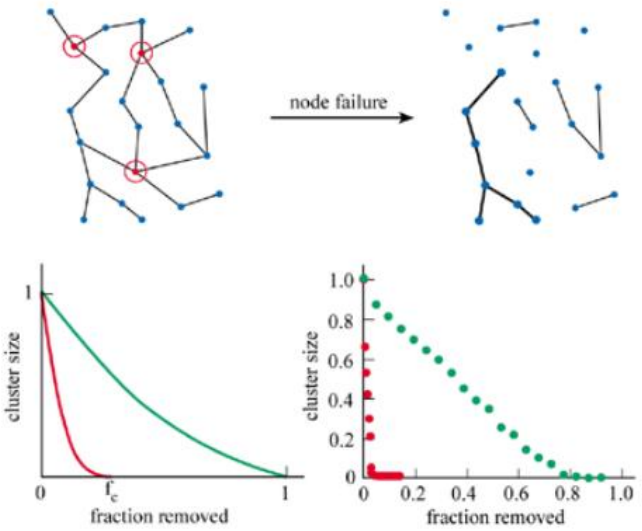

Figura 1: Fragmentación de una red libre de escala.

2.4 Detección de comunidades

Una rama del estudio de las redes complejas es la detección de comunidades. Se basa en la observación de que multitud de redes están internamente fragmentadas en grupos de distintos nodos, o módulos. La identificación de estos grupos o módulos ayuda a comprender de una forma más completa la configuración o funcionalidad de una red. Sin embargo, la detección de comunidades no es tarea simple, ya que cada comunidad puede tener su propio tamaño o densidad o incluso un nodo puede solaparse entre varias comunidades, es decir, que pertenezca a varias comunidades a la vez. Sin embargo se han desarrollado varios métodos de detección de comunidades, cada uno con sus ventajas e inconvenientes. En el presente trabajo se va a hacer uso de estos métodos para el análisis del comportamiento u opinión política sobre grafos obtenidos a partir de datos extraídos de Twitter en el marco de una campaña electoral.

2.5 Modularidad

La modularidad es una forma de evaluar la estructura de una red, midiendo el nivel de fuerza de la división de una red en agrupaciones o comunidades (Newman & Girvan, 2004). Una red con un alto índice de modularidad significa que en esa red los nodos de una misma comunidad están fuertemente enlazados mientras que los nodos de distintas comunidades están débilmente enlazados. Es una medida utilizada para la detección de comunidades en redes de grafos.

La modularidad expresa la fracción de aristas dentro de una misma comunidad frente al valor esperado en una red aleatoria. Por lo que si la fracción de aristas dentro de las

comunidades es similar a la aleatorización el valor de la modularidad se acercaría a 0. Por otro lado el valor se acercaría a 1 en el caso de que la fracción de aristas dentro de los mismos grupos sea superior al grafo aleatorio.

Definición: Sea una red de v vértices y m enlaces, en la que cada vértice tiene un grado k_v, el número total de posibles reconexiones entre enlaces sería:

$$l_n = \sum_{v}^{n} k_v = 2m \quad (1)$$

El número esperado (aleatorio) de enlaces quedaría como (enlaces completos entre los nodos v y w) / (número total de posibles reconexiones):

$$k_v * k_w / l_n \; = \; k_v * k_w / 2m \quad (2)$$

Por lo tanto el número real de enlaces entre v y w (la matriz de adyacencia A_{vw}) menos el número esperado (la fórmula 2) sería:

$$A_{vw} - \frac{k_v * k_w}{2m} \quad (3)$$

La ecuación de la modularidad quedaría como (4) con c_v como la comunidad a la que pertenece v, donde $\delta(c_v, c_w)$ es igual a 1 si $c_v = c_w$ y 0 en caso contrario.

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v * k_w}{2m} \right] \delta(c_v, c_w) \quad (4)$$

Usando $\delta(cv, cw) = \sum_i \delta(cv, i)\delta(cw, i)$, se puede definir e_{ij} como la fracción de los enlaces que unen vértices de la comunidad i a la comunidad j

$$e_{ij} = \frac{1}{2m} \sum_{vw} A_{vw} \delta(c_v, i) \delta(c_w, j) \quad (5)$$

y a_i como la fracción de enlaces que salen de la comunidad i

$$a_i = \frac{1}{2m} \sum_{v} k_v \delta(c_v, i) \quad (6)$$

Finalmente se puede obtener la expresión simplificada de modularidad a través de las ecuaciones (5) y (6):

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v * k_w}{2m} \right] \delta(c_v, c_w) = \sum_{i=1}^{c} (e_{ii} - {a_i}^2) \quad (7)$$

Esta medida es válida para grafos no dirigidos y sin pesos en sus enlaces. Si el grafo es ponderado la forma sería la siguiente, siendo s_i la suma de los pesos de los enlaces que

van al nodo i, w es la suma de los pesos de todos los enlaces del grafo y W la matriz simétrica que contiene los pesos de los enlaces que unen a los nodos v y w.

$$Q = \frac{1}{2w}\sum_{vw}\left[W_{vw} - \frac{s_v * s_w}{2w}\right]\delta(c_v, c_w) \quad (8)$$

Si el grafo es dirigido la formula cambiaría ligeramente, teniendo en cuenta el grado exterior del vértice v y el grado interior del vértice w. Siendo la suma de grados interiores y exteriores m.

$$Q = \frac{1}{m}\sum_{vw}\left[A_{vw} - \frac{k_v^{out} * k_w^{in}}{m}\right]\delta(c_v, c_w) \quad (9)$$

Finalmente, si el grafo es combinación de dirigido y ponderado, la formula quedaría de la siguiente manera:

$$Q = \frac{1}{w}\sum_{vw}\left[W_{vw} - \frac{s_v^{out} * s_w^{in}}{w}\right]\delta(c_v, c_w) \quad (10)$$

2.6 Métodos de detección de comunidades

Actualmente existen múltiples enfoques para la detección de comunidades en grafos. A continuación se van a explicar los más extendidos.

2.6.1 Clustering jerárquico

Este método se basa en el cálculo de similitudes o distancias entre los nodos del grafo (Jaccard, Hamming, Euclídea, coseno, etc). Según el orden en el que se forman los subgrupos tiene dos planteamientos: aglomerativo o divisivo. En el tipo aglomerativo se parte de los nodos individualizados y en cada iteración se le añaden clústeres hasta finalmente obtener un único clúster. En cambio el tipo divisivo es el inverso, se parte de un único clúster hasta obtener tantos grupos como nodos tiene el grafo. Los resultados obtenidos por los métodos de clustering se representan mediante los llamados dendogramas (ver Figura 2).

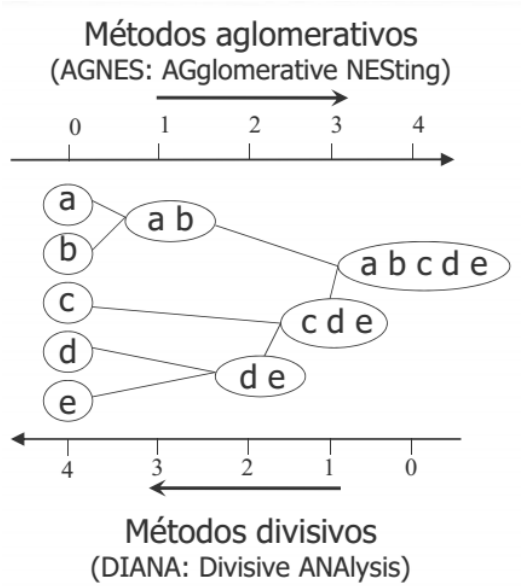

Figura 2: Métodos de clustering, de izquierda a derecha aglomerativo, de derecha a izquierda divisivo.

Estas técnicas por sí mismas no ofrecen una manera de saber en qué momento se obtiene la partición óptima, por lo que se suelen basar en heurísticas. Una estrategia desde el enfoque divisivo muy popular es la desarrollada por Girvan-Newman (Girvan & Newman, 2002), basada en la intermediación o betweennesses de los enlaces (número de caminos más cortos que pasan por un enlace). El procedimiento de esta estrategia es el siguiente:

1. La intermediación de todos los enlaces existentes en la red se calcula primero.
2. Se elimina el enlace con la más alta intermediación.
3. La intermediación de todos los enlaces afectados por la eliminación se vuelve a calcular.
4. Pasos 2 y 3 se repiten hasta que no hay quedan enlaces.

Sin embargo este método tiene ciertas desventajas. La eliminación de los enlaces puede repercutir en el betweennesses del resto, con lo que hay que recalcularlo para todos, lo que hace que el algoritmo no escale bien a medida que el grafo se vaya haciendo más grande.

2.6.2 Optimización de la modularidad

Este enfoque trata de maximizar la medida de modularidad. Se basa en aplicar heurísticas, ya que la búsqueda exhaustiva de particiones que optimicen esta medida es irresoluble, siendo las más populares las técnicas voraces (greedy). Un ejemplo de algoritmo basado en una estrategia greedy es Louvain (recibe el nombre de la universidad de Louvain) creado por Blondel y otros (Blondel et al., 2008), cuyo funcionamiento resumido es el siguiente:

1. Comenzar con todos los nodos.
2. Unir iterativamente los clústeres que impliquen un mayor incremento ΔQ de modularidad.
3. Parar cuando dicho incremento sea menor que 0.

No obstante, la optimización de la modularidad tiene un límite de resolución, lo que significa que no puede detectar subgrupos de tamaño menor que un umbral determinado por el número de enlaces y patrón de conexiones del grafo(Fortunato & Barthélemy, 2007).

2.6.3 Caminos aleatorios

Este método se basa en la simulación de cómo se difunde la información a través de una red, mediante el uso de caminos aleatorios. Se fundamenta en que un caminante aleatorio al encontrarse con una zona de la red con una alta concentración de enlaces quedaría atrapado dentro de la misma, descubriendo una nueva comunidad. Este procedimiento se modela a través de cadenas de Markov, en las que cada estado de la cadena es un nodo con una determinada probabilidad de cambiar de estado (visitar otro nodo adyacente). La probabilidad de transición entre el nodo v y el nodo w es $P_{vw} = \frac{A_{vw}}{k_w}$, siendo A_{vw} la matriz de adyacencia y k_w el grado de w. Por lo tanto la probabilidad de ir de v a w mediante un camino de longitud l sería P_{vw}^l.

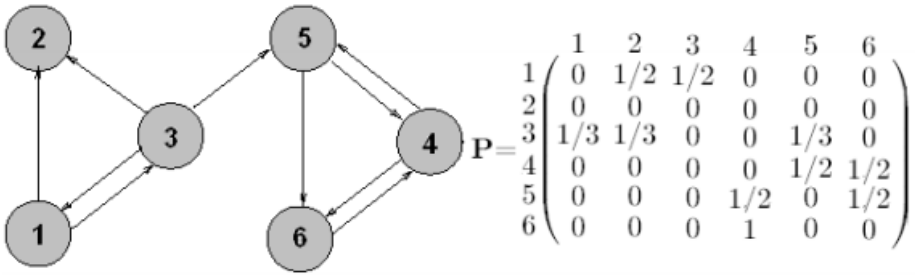

Figura 3: Ejemplo de matriz de transición (derecha) para un grafo (izquierda)

Combinando este concepto con los sistemas de clustering aglomerativo se obtiene un método que permite encontrar comunidades a diferentes escalas, siendo el más conocido el denominado Walktrap (Pons & Latapy, 2005). En este caso la métrica de similitud estaría determinada en base a la distancia entra comunidades, definida por la siguiente ecuación, siendo $P_{C_1 v}^l$ la probabilidad de ir desde una comunidad a un nodo v mediante un camino de longitud l:

$$r_{C_1 C_2} = \sqrt{\sum_v^n \frac{(P_{C_1 v}^l - P_{C_2 v}^l)^2}{k_v}} \quad (11)$$

3 Diseño y Desarrollo

3.1 Medios técnicos empleados

Los medios técnicos empelados para nuestro estudio como lenguajes de programación o librerías han sido diversos. Para el desarrollo de la aplicación encargada de la obtención de datos de Twitter se utilizó el lenguaje Java junto con la API Twitter4j (no oficial) así como una conexión con una base de datos Sqlite. Por otro lado, para la representación gráfica de los grafos se utilizó la herramienta de visualización Gephi. En cuanto al procesado de los datos, análisis y estudio de los grafos obtenidos se utilizó el lenguaje R junto con la librería igraph (Patrick A. et al., 2015). Finalmente, para el estudio de la polaridad de los tweets a través de clasificadores se utilizó la librería de R pscl (Political Science Computational Laboratory) (Jackman S. et al., 2008) centrada en el análisis político.

3.2 Recolección de datos y metodología

Para realizar una aplicación que se valga del API de Twitter, hay que acceder a la página de desarrolladores de Twitter con usuario y contraseña de la red social. Una vez registrado hay que dar de alta la aplicación, obteniendo una serie de claves ("Consumer Key", "Consumer secret", etc.) mediante las cuales se realiza la conexión con Twitter desde el programa.

La aplicación desarrollada tiene como objetivo la recolección de datos de Twitter durante la campaña electoral para su futuro análisis. Los datos se obtienen a través de búsquedas prefijadas mediante parámetros de entrada. Los parámetros escogidos para la búsqueda son los hashtags (la aplicación no es sensible a las mayúsculas) sobre los que se filtrarán los tweets a recolectar. En el caso de que se introduzcan varios parámetros**,** la aplicación realiza un or (disyunción) entre los hashtags a la hora de filtrar los tweets. Asimismo, el método de recolección utilizado ha sido diseñado basado en la Streaming API de Twitter, que proporciona un subconjunto del flujo total de tweets en tiempo real y que a diferencia del método RESTful (basado en peticiones a los servidores de Twitter), no tiene limitaciones de peticiones por hora o usuario. El procedimiento de la aplicación es el siguiente: la descarga de tweets la hemos orientado a búsquedas por palabras clave (los hashtags). Dada una búsqueda, Twitter filtra los tweets que contengan alguno de las palabras clave, devolviendo a través del API una lista de tweets. Dado un tweet, nuestra aplicación comprueba si hace referencia a otro tweet a través de alguna relación (retweet, mención o reply), en tal caso se descarga el tweet referenciado, y así recursivamente con cada tweet dado.

Los hashtags utilizados para la extracción de datos de nuestro estudio durante las elecciones de la comunidad de Cataluña fueron “27S” y “eleccionescatalanas”. Para las elecciones generales se han utilizado “20D” y los hashtags de los partidos “PSOE”, “Ciudadanos”, “PartidoPopular”, “Podemos”, “IzquierdaUnida” y “UpyD”. Estos hashtags han sido seleccionados para nuestro estudio a raíz de la noticia de que Twitter habilitaría, de cara a las elecciones, emojis con el logotipo de cada partido en caso de escribir el hashtags de uno de los seis principales partidos políticos. La empresa Twitter tenía intención de convertir a la red social en el principal soporte para seguir la actualidad electoral (Twitter España, 2015). El criterio de selección planteado ha sido la suposición de que la aparición del logotipo de los partidos en los tweets, puede favorecer el envío de tweets de defensa o apoyo a partidos, al permitirse hacer esta comunicación algo más gráfica con los emojis.

Los datos se obtuvieron durante los periodos del 24 al 27 de Septiembre con ciclos de 24 horas en el caso de las elecciones de la comunidad de Cataluña y del 14 al 20 de Diciembre con ciclos de 16 horas (entre las 8:00 y las 24:00) para las elecciones generales. En este último caso no se pudieron repetir las mismas condiciones que en las elecciones catalanas debido a limitaciones en la disponibilidad, por lo que no se debe interpretar la posible diferencia de volumen de datos como un aumento o mejora de la participación en Twitter de unas elecciones a las siguientes. La Tabla 1 muestra una descripción de los datos en cada uno de los escenarios analizados.

Datos	Usuarios	Tweets	Retweets	Menciones	Total Tweets	Hashtags
Elecciones Catalanas	131774	89732	302424	16327	408483	11116
Elecciones Generales	374098	391590	714640	58454	1164684	27613

Tabla 1: Información de los conjuntos datos: número de usuarios, número de tweets, número de retweets, número de menciones, total de tweets (tweet + retweet + mención) y número de hashtags.

El funcionamiento interno de la aplicación a la hora de interactuar con la base de datos consiste en: crear la base de datos en caso de no existir, descargar los tweets en tiempo real a través de la utilización de un listener (Status Listener) suministrado por la API de Twitter y finalmente la introducción de tweets, usuarios y hashtags en la base de datos de forma recursiva en caso de encontrarse otros tweets referenciados mediante relación de quote, retweet, etc.

La estructura de la base de datos es similar a las clases proporcionadas por la API Twitter4j, modelándose mediante tres tablas:

- La tabla Tweets guarda el texto del tweet en cuestión, la fecha de creación, información geográfica del lugar donde se realizó (latitud, longitud y país), si se trata de un retweet/quote/reply, el número de retweets y de favoritos y una referencia al usuario que lo generó.

- Users almacena información de usuario, como su nombre, alias, nacionalidad, lenguaje, número de seguidores (followers), si se trata una cuenta verificada, etc.
- Hashtags sirve para referenciar un hashtag con un determinado tweet, no se ha modelado como campo de la tabla Tweets puesto que un tweet puede contener múltiples hashtags (de hecho un fenómeno frecuente es la creación de tweets sin texto escrito, únicamente formados por hashtags).

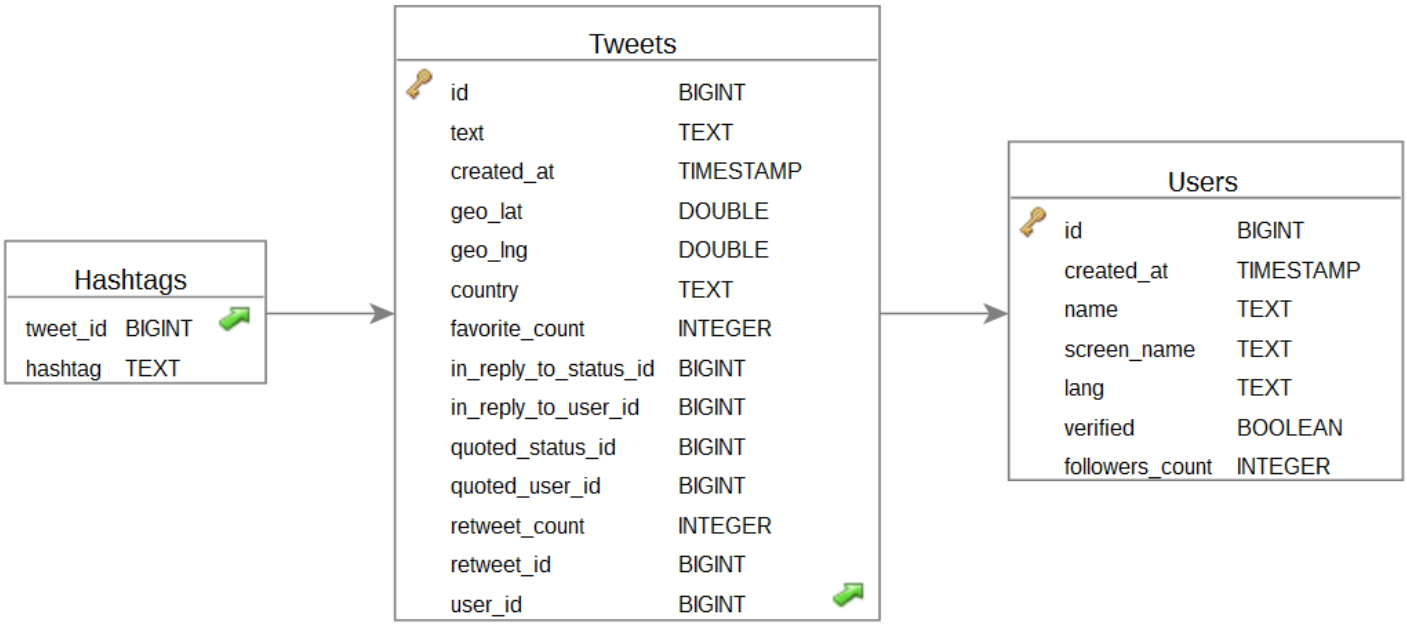

Figura 4: Estructura de la base de datos SQLite (las primary keys aparecen con el icono de una llave, los campos que referencian a otras tablas aparecen con una flecha verde).

4 Resultados

4.1 Topología y métricas de los grafos

Hay investigadores que ven la relación de retweet como respaldo (*endorsement*) de la información, puesto que el uso repetido los RTs es un indicador de la opinión política (Conover et al., 2012). Lo cierto es que existe mucha polémica en torno a este enfoque, ya que no toma en cuenta el uso irónico del RT, llegando a poder tener consecuencias judiciales. En el año 2015 el FBI arrestó a un joven de 17 años de edad utilizando como evidencias el haber realizado regularmente retweets de las declaraciones del líder del Daesh, por lo que se deduce que jurídicamente la relación retweet supone respaldo (K. Knibbs, 2015). Por otro lado, el jefe de redacción del New York Times se posicionó a favor del *"RTs ≠ endorsements"* (retweets no son respaldo): *"en general, creo que los usuarios de Twitter por ahora entienden que un retweet implica compartir o que señala algo, no necesariamente defender o apoyar [..]"* (Kirkland, 2014).

Sin embargo parece aceptada la interpretación del retweet como indicio de respaldo desde un punto de vista estadístico, y así en este trabajo hemos considerado tomar como relaciones de respaldo tanto el retweet como el uso de hashtags y las menciones (quotes). De esta forma los grafos construidos son tres, uno por cada tipo de relación, cuyos enlaces serán dirigidos. En las relaciones de retweet y quote, el peso de estos enlaces será establecido por el número de veces que un usuario ha citado o realizado retweet a otro usuario. En cambio, en el caso del grafo generado a través de la relación de hashtag, el enlace une a un usuario u y a un hashtag h, con peso w si el usuario u ha utilizado w veces el hashtag h en sus tweets. Esta configuración da al grafo con relación hashtag la apariencia de un cristal de nieve, es decir, en el grafo conviven dos tipos de nodos: los usuarios y los hashtags. Además, los únicos enlaces existentes son entre usuario y hashtag, quedando los usuarios "colgando" de los hashtag.

Tras obtener los grafos resultantes de estos tres tipos de relaciones, calculamos diversos datos descriptivos de los mismos. Las propiedades de los grafos analizadas han sido las siguientes:

- Grado Medio: se define el grado medio del grafo como la media del grado de los nodos. El grado de un nodo es el número de enlaces conectados a ese nodo.
- Diámetro: es la distancia de grafo más larga entre dos nodos cualesquiera en la red. Representa el tamaño del grafo y permite saber lo grande que es el mismo.
- Longitud media de camino: consiste en la distancia media entre todos los pares de nodos de la red. Una red muy densa tenderá a tener una longitud media menor, puesto que existen muchos más caminos para llegar de un nodo a otro.

- Reciprocidad: esta propiedad consiste en la probabilidad de que si existe una arista en un sentido también exista una en el sentido opuesto.
- Densidad: esta propiedad mide la conectividad de una determinada red. Es el cociente entre el número de enlaces de la red y el número de enlaces posibles. Un grafo completo, es decir, aquel grafo que tenga todos los enlaces posibles tiene una densidad igual a 1. En cambio, se denomina grafo disperso a aquel que tiene un valor de densidad pequeño.
- Componentes fuertemente conexas: un grupo de nodos es un componente fuertemente conexo si para cada par de nodos *u* y *v* existe un camino de *u* hacia *v* y un camino de *v* hacia *u*.
- Coeficiente medio de clustering: esta propiedad consiste en la media de los coeficientes de clustering de cada nodo del grafo. El coeficiente de clustering o agrupamiento de un nodo en un grafo cuantifica qué tanto está de agrupado (o interconectado) con sus vecinos.

La Tabla 2 muestra un resumen de cada una de los grafos obtenidos así como sus propiedades generales.

Elecciones de Cataluña					
Relación	Nodos	Enlaces	Grado Medio	Diámetro	Long. Media de camino
Retweet	102004	254099	4.98	24	7.00
Hashtag	68902	202379	5.87	2	1.00
Quote	12461	14635	2.34	12	3.44
Elecciones generales					
Relación	Nodos	Enlaces	Grado Medio	Diámetro	Long. Media de camino
Retweet	234643	578270	4.92	21	6.25
Hashtag	161739	474346	5.86	1	1.00
Quote	41664	49637	2.38	34	13.07

Tabla 2: Resumen de las propiedades generales de cada uno de los grafos generados.

El diámetro y las longitudes medias de camino en los grafos generados son relativamente pequeños respecto al número de enlaces y nodos de los mismos. Este fenómeno se conoce comúnmente como red de mundo pequeño. Este tipo de grafo es aquel en el que no todos los nodos son vecinos entre sí, pero en el cual se puede llegar a un nodo desde cualquier nodo a través de un corto número de saltos. Este fenómeno se observa en diferentes grafos, como pueden ser los generados en redes sociales, enlaces de páginas web, etc. (Barabási, 2003).

En la Tabla 3 se puede observar que los valores para la propiedad de reciprocidad son muy pequeños en todos los tipos de grafos generados. Esto ocurre a causa de la configuración de las relaciones dentro de la red social Twitter. Las relaciones suelen estar orientadas en una sola dirección, es decir, un usuario puede hacer a otro retweet, mención, etc. pero no necesariamente al revés.

En cuanto a la propiedad de densidad, los grafos obtenidos unos valores tienen muy bajos, lo que los convierte en grafos de tipo disperso. Respecto al coeficiente medio de clustering se dice, que si este es significativamente más alto que un grafo aleatorio construido con el mismo conjunto de vértices y si el grafo tiene aproximadamente la misma longitud media de camino más corto que su correspondiente grafo aleatorio, el grafo analizado se podría considerar de mundo pequeño.

Elecciones de Cataluña						
Relación	Componentes fuertemente conexas	Componentes débilmente conexas	Coeficiente medio de clustering	Componente conexa más grande	Densidad	Reciprocidad
Retweet	98573	1502	0.09	98411 nodos	2.44e-05	0.007
Hashtag	68902	1039	0	66231 nodos	4.26e-05	0
Quote	12434	1149	0.02	9890 nodos	9.42e-05	0.002
Elecciones Generales						
Relación	Componentes fuertemente conexas	Componentes débilmente conexas	Coeficiente medio de clustering	Componente conexa más grande	Densidad	Reciprocidad
Retweet	220897	9598	0.10	210965 nodos	1.05e-05	0.01
Hashtag	161739	4796	0	148029 nodos	1.81e-05	0
Quote	40685	5812	0.02	28263 nodos	2.85e-05	0.02

Tabla 3: Resumen de las propiedades de conectividad de cada uno de los grafos generados.

4.1.1 Distribución del grado de los nodos

La Figura 5 y la Figura 6 evidencian que todas las redes de grafos generadas con cada una de las relaciones (Retweet, Hashtag y Quote) son libres de escala ya que muestra una distribución de grado de ley de potencia, que al ser normalizada a escala logarítmica se aproxima a una recta descendente. La característica que da nombre a este tipo de redes (libres de escala) es que mientras que en otras redes la función de distribución de grados tiene un máximo en el valor medio del grado, es decir hay un valor característico, en las redes libres de escala no lo hay, no son homogéneas, la distribución de su grado no se concentra en torno una escala.

Como se puede observar en las gráficas, alrededor del 20% de los nodos de cada grafo concentra sobre el 80% del grado, mientras que por el contrario, el 80% de los nodos atesoran el 20% del grado. Asimismo se observa una alta densidad de nodos tanto en las gráficas de Retweet como de Hashtag, frente a la gráfica de Quote. Esto es así debido a la considerable diferencia de tamaño entre los grafos.

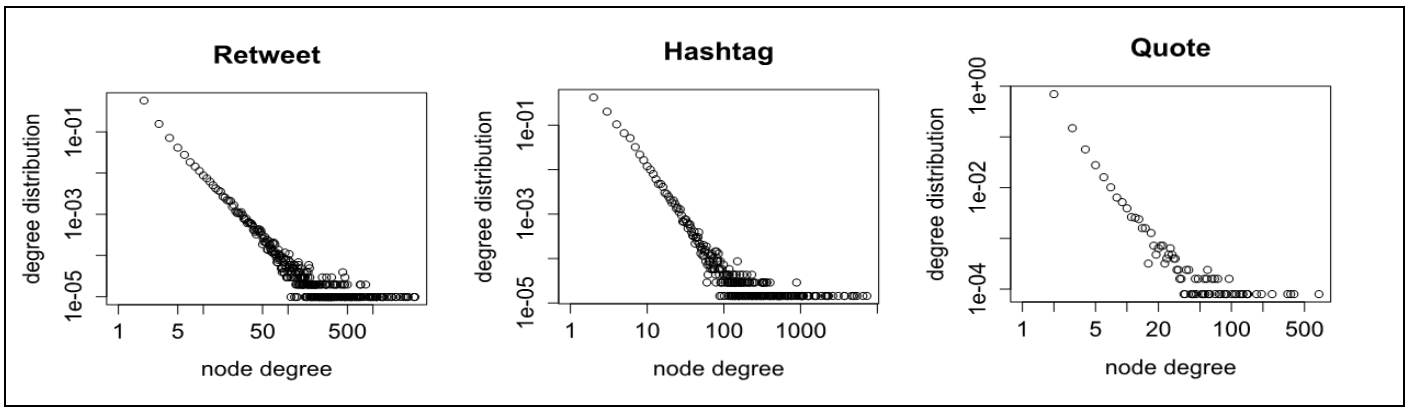

Figura 5: Distribución del grado de los nodos sobre el conjunto de datos de las elecciones de Cataluña del año 2015. Las gráficas están en escala logarítmica.

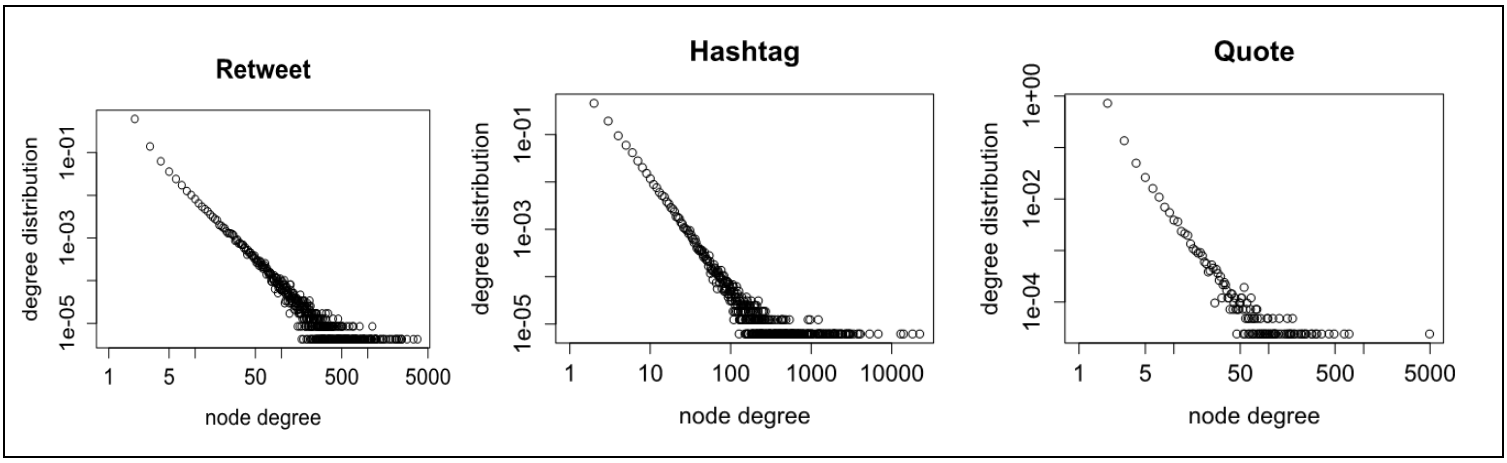

Figura 6: Distribución del grado de los nodos sobre el conjunto de datos de las elecciones generales del año 2015. Las gráficas están en escala logarítmica.

4.2 Descripción de las comunidades

La detección de comunidades se realizó mediante el método Louvain. La Figura 7 y la Figura 8 muestran el resultado de aplicar el algoritmo en los grafos obtenidos mediante las relaciones de Retweet en las dos campañas electorales analizadas. El resto de figuras se puede consultar en el anexo C. Sobre cada uno de los grafos se ha aplicado una selección de las comunidades consideradas como representativas de partidos políticos.

El color de los nodos corresponde a las comunidades de partidos políticos detectadas por el algoritmo y el de los enlaces al color de la comunidad que parten. Así, en el marco de las elecciones catalanas, la configuración de colores sería la siguiente: PP (azul), PSOE (rojo), Catsiqueespot (granate), C's (naranja), JxSi (turquesa), y CUP (amarillo). En el contexto de las elecciones generales se mantendrán los colores excepto UPYD (rosa) y Unida Popular (verde) y PODEMOS (morado).

La caracterización política de las comunidades detectadas se realizó mediante un etiquetado manual (ground truth), escogiendo los nodos representativos de carácter político (la cuenta de un candidato, de un partido o en el caso de ser relación de hashtags, el hashtag de la campaña de un determinado partido, etc) con mayor grado dentro de la misma comunidad. La Tabla 4 y la Tabla 5 muestran estas cuentas así como la comunidad a la que pertenecen y el tamaño de las mismas por cada escenario.

Se ha escogido el método Louvain para realizar nuestro estudio debido a que las comunidades que genera frente a las resultantes de aplicar Walktrap o Infomap son más densas en comparación. Además, han sido en las comunidades formadas a través de Louvain donde la polaridad política se ha encontrado de una forma más evidente, debido que al realizar el etiquetado manual se observaba una mayor concentración de representantes políticos en ciertas comunidades, cosa que mediante los otros algoritmos no era tan obvio.

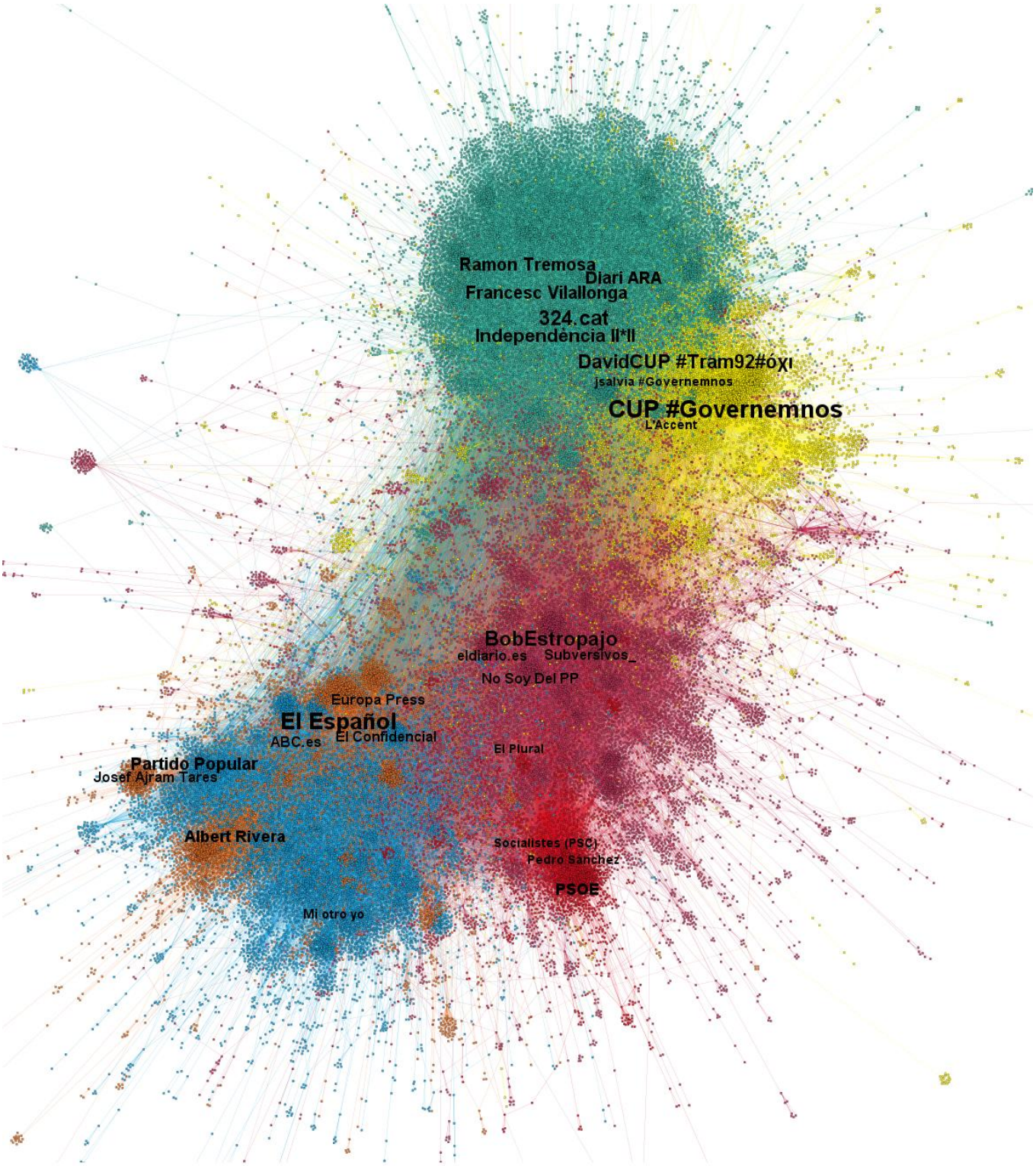

Figura 7: Grafo de Retweets de las elecciones catalanas del 2015
(en negrita los usuarios con mayor grado).

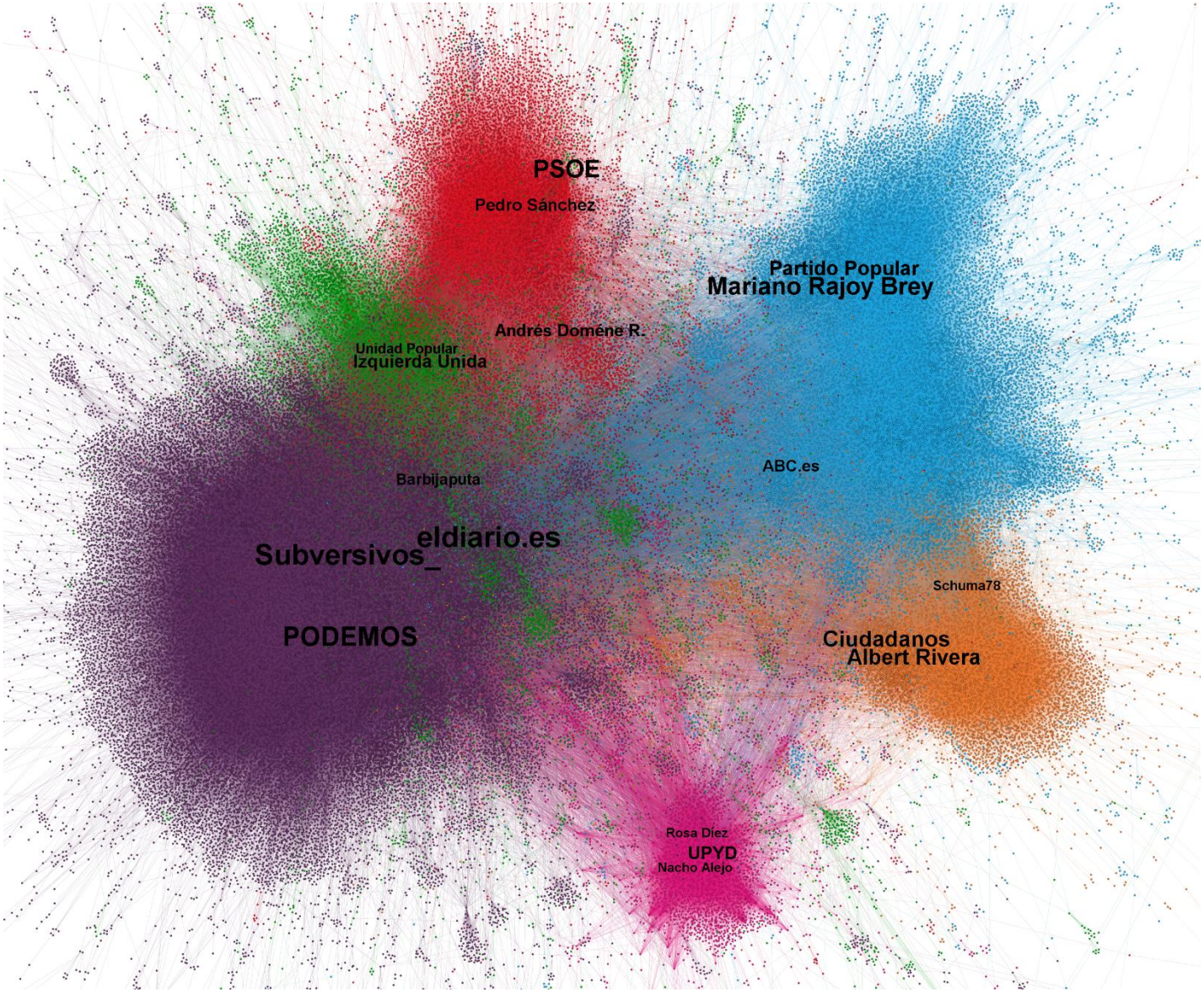

Figura 8: Grafo de Retweets de las elecciones generales del 2015 (en negrita los usuarios con mayor grado). .

Elecciones de Cataluña									
	Relación Retweet			Relación Quote			Relación Hashtag		
Partido	**ID**	**Cuentas**	**Tamaño**	**ID**	**Cuentas**	**Tamaño**	**ID**	**Cuentas**	**Tamaño**
C's	67	Ciudadanos, Albert Rivera, Ciudadanos Madrid, C's Cornellà	7233	83	Ciudadanos, Ciudadanos Marbella, Albert Rivera	197	99	#yovotonaranja, #apoderadoscs, #bcnnaranja, #mataronaranja	1965
Catsiq	47	PODEMOS, Ada Colau, PODEM, CAT Sí que es Pot	13701	140	PODEMOS, CAT Sí que es pot, PODEM	193	305	#catsiqueespot27s, #votacatsiqueespot, #catalunyasiqueespot, #catsiqueespot	3007
CUP	10	CUP #Governemnos, DavidCUP #Tram92#όχι, Antonio Baños, CUP Barcelona	6709	173	CUP #Governemnos, DavidCUP #Tram92#όχι, 27S #CUP, CUP Arenys de Munt	319	248	#governemnos, #cup, #somriurecup, #sindicalistesperlacup	4877
JxSi	14	Junts pel Sí, Assemblea Nacional, Independència ll*ll	19696	119	Junts pel Sí, Assemblea Nacional, Independència ll*ll	507	13	#juntspelsi, #votapermi, #independencia, #presidentmas	9099
PP	413	Mariano Rajoy Brey, Partido Popular, PP de Madrid, Xavier García Albiol	11241	195	Partido Popular, PP Català, Pablo Casado Blanco	441	61	#plantemoscara, #cataluñaespaña, #unidosganamos, #plantemcara	5620
PSC	19	PSOE, Pedro Sánchez, Socialistes (PSC), PSC Barcelona	3058	91	PSOE, Pedro Sánchez, Socialistes (PSC)	277	10	#votasocialsita, #tenimsolucions, #iceta27s, #votaiceta	2120

Tabla 4: Comunidades de partidarios encontradas en cada grafo generado (relación retweet, quote y hashtag) a través de los datos de las elecciones catalanas junto con las cuentas más significantes con más grado y su tamaño.

Elecciones Generales									
	Comunidad Retweet			Comunidad Quote			Comunidad Hashtag		
Partido	**ID**	**Cuentas**	**Tamaño**	**ID**	**Cuentas**	**Tamaño**	**ID**	**Cuentas**	**Tamaño**
C's	2443	Ciudadanos, Albert Rivera, Ciudadanos Madrid, Ciudadanos Valencia	7118	152	Ciudadanos, Albert Rivera, Ciudadanos Madrid, Ciudadanos Andalucía	750	3	#ciudadanos, #venceralailusion, #jovenesconalbert, #albertpresidente	10974
POD	177	PODEMOS, Pablo Echenique, Íñigo Errejón, Pablo Iglesias	37058	335	PODEMOS, Pablo Echenique, Íñigo Errejón, Pablo Iglesias, En Comú Podem	3070	4019	#podemos, #sisepuede, #podemos20d, #votapodemos20d	22109
PP	411	Mariano Rajoy Brey, Partido Popular, Cristina Cifuentes, Sáenz de Santamaría	19786	279	Mariano Rajoy Brey, Partido Popular, Mª Dolores Cospedal, Cristina Cifuentes	3268	5	#partidopopular, #votapp, #yovotopp, #vamospp	12510
PSOE	565	PSOE, Pedro Sánchez, Susana Díaz, PSOE de Andalucía	8564	83	PSOE, Pedro Sánchez, PSOE Congreso, PSOE Albacete	1176	7	#psoe, #votapsoe, #pedropresidente, #rojopsoe	10239
UP	1153	Izquierda Unida, Unidad Popular, PCAndalucía, Alberto Garzón Blanc	7790	1700	Izquierda Unida, Izquierda Castellana, IU Avila	373	1292	#unidadpopular, #izquierdaunida, #yovotoup, #yoagarzon	3905
UPYD	927	UPYD, Rosa Díez, Andrés Herzog, UPyD Madrid	3074	191	UPYD, Rosa Díez, Andrés Herzog, UPyD Berlín	1001	406	#upyd, #votaupyd, #apoderadosupyd, #masespaña	3470

Tabla 5: Comunidades de partidarios encontradas en cada grafo generado (relación retweet, quote y hashtag) a través de los datos de las elecciones generales junto con las cuentas más significantes con más grado y su tamaño.

4.3 Vínculos entre las comunidades de partidarios

Siguiendo las investigaciones de comunidades de partidarios del profesor Esteban Moro, se ha realizado un análisis de las relaciones entre las comunidades detectadas (en concreto sobre la relación de RTs). Para obtener los flujos significativos de comunicación entre comunidades se ha utilizado el método original empleado por el autor. Su aplicación ha consistido en realizar una comparación del número de enlaces (retweets) entre comunidades con el número de enlaces que se hubiera repartido aleatoriamente, mostrando únicamente aquellos cuyo z-score es mayor que 2 respecto a la simulación aleatoria (Moro et al., 2014).

El enfoque de Moro consistió en observar la polarización de las comunicaciones entre comunidades afines durante las campañas electorales de la comunidad de Cataluña de 2010 y 2012. Su estudio confirmó que las relaciones se encontraban más concentradas entre grupos de comunidades nacionalistas, de izquierdas, de centro-derecha, etc., viendo una estructura jerárquica de las relaciones tanto a pequeña escala (dentro de una comunidad), como a escalas más grandes (entre comunidades con una misma ideología política) (Moro et al., 2014).

Los resultados obtenidos en nuestro estudio dentro del marco de las elecciones catalanas siguen el mismo patrón observado con anterioridad: concentración de las comunicaciones entre los grupos nacionalistas (JxSi y CUP) junto con los abiertos a un referéndum (Catsiqueespot) y entre los grupos anti-independentistas (PP y C's). La cuestión es que esta organización de las relaciones no está conformada por una afinidad ideológica, sino por una postura política concreta: la posible independencia de Cataluña. Este acontecimiento está provocando *"extraños compañeros de cama"*, tanto en las redes sociales como fuera de ellas: partidos como Convergencia (dentro de la plataforma JxSi) y la CUP, que ideológicamente están en las *antípodas*, presentan fuertes vínculos en sus relaciones.

Sin embargo, los resultados obtenidos durante las elecciones generales se desvían ligeramente de este patrón de afinidad. Existe una organización grupal ideológica, grupos de centro-izquierda (PSOE, PODEMOS y Unidad Popular) y de derecha-centro (PP y C's), pero la agrupación de PP y PODEMOS tienen uno de los enlaces más fuertes, siendo ideológicamente opuestos. Si se observan los gráficos (ver Figura 9) se puede ver como el gráfico de las elecciones generales gira en torno a la comunidad de PODEMOS. Pero, ¿cuál es el por qué se da este suceso? Esto puede ser ocasionado debido a que el partido de PODEMOS es uno de los que más presencia tienen en las redes sociales. En 2015 contaban con entre 20 y 25 *community managers* voluntarios para la gestión de su cuenta en Twitter y ganaron más de 100.000 seguidores ese mismo año (Lorente, 2015). Otra razón sería el reflejo de su irrupción en el mapa político y mediático, lo que desestabilizaría la postura que toma el retweet como respaldo, por lo menos en los casos de surgimiento de fuerzas políticas con alta presencia mediática, siendo algo a tomar en cuenta en posteriores investigaciones.

Por estas razones en este documento se ha preferido utilizar el término vínculo en lugar de afinidad. Los datos observados han puesto en manifiesto que estas relaciones no siempre se dan por causa de una afinidad ideológica, aunque habría que reconocer que los marcos en los que se han estudiado estas relaciones no son políticamente estables (independencia de Cataluña y surgimiento de fuerzas por *"el cambio"*).

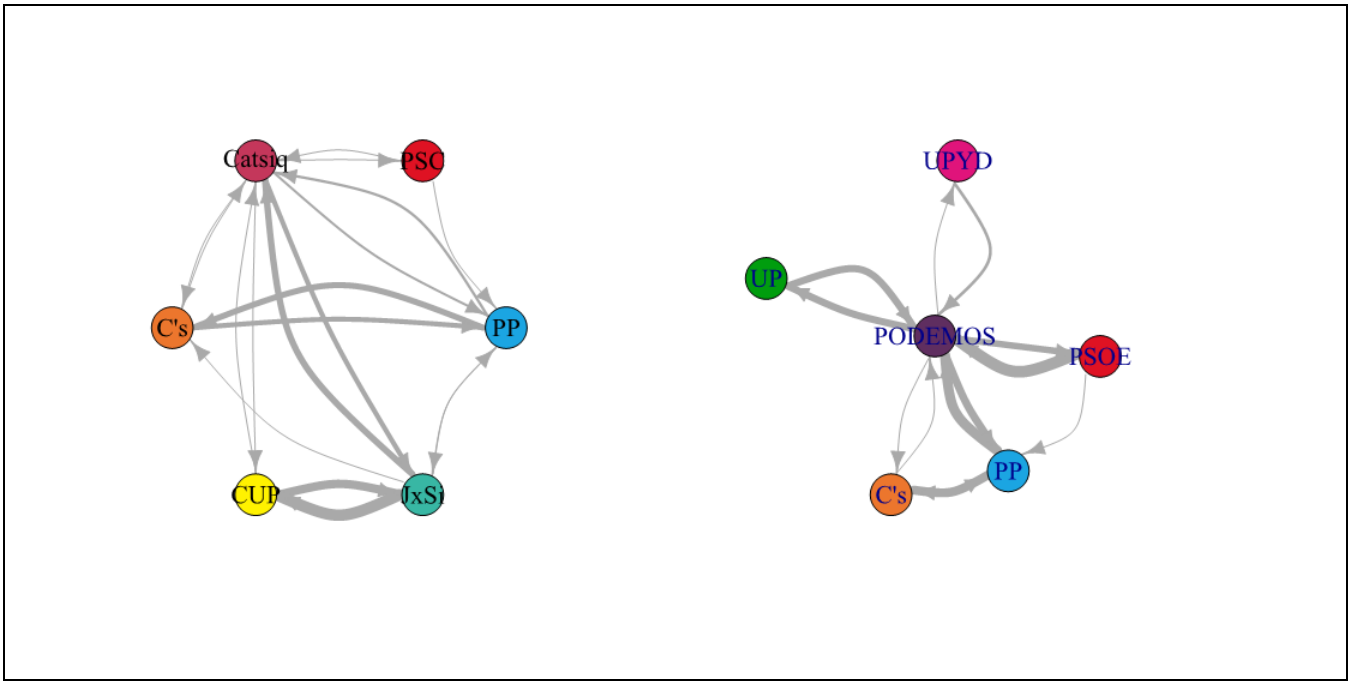

Figura 9: Grafos simplificados de la comunicación (RTs) entre las comunidades detectadas por el algoritmo durante las elecciones de Cataluña de 2015 (izquierda) y las elecciones generales de 2015 (derecha). La anchura de los enlaces es proporcional al número de RTs.

4.4 Intención de voto y comunidades

Tras haber observado la forma polarizada en que se realizan las comunicaciones entre comunidades, se ha pasado a analizar la posibilidad de correlación entre los tamaños de las mismas y el porcentaje de votos obtenido. Siguiendo el enfoque del profesor Moro, el estudio se ha realizado sobre las comunidades generadas por la relación de Retweet y ampliando el análisis con las relaciones de Hashtag y Quote. En cuanto al análisis del número de menciones como índice de voto, también se ha realizado una ampliación en los estudios, combinándolo con dos factores.

El primer factor consiste en el número de seguidores (*followers*) de cada partido. El número de seguidores da un indicio de impacto (cuántos usuarios leen los posts) que puede tener la información que publica un usuario. Como muestra de la relevancia de esta dato, en el 74% de las campañas electorales analizadas en EEUU, los candidatos con más seguidores alcanzaron mejores posiciones finales (Zarrella, 2010). El segundo factor se basa en el

sentimiento positivo de los tweets de cada partido. Para ello se ha utilizado la herramienta LIWC2015, analizando los tweets del conjunto de datos que contenían exclusivamente los hashtag de cada uno de los seis principales partidos para las elecciones generales y los hashtags de campaña para las elecciones catalanas (#ciudadanos, #votasocialista, #juntspelsi, #governemnos, #catsiqueespot27s y #plantemoscara). Las principales razones por las que se ha hecho uso de esta herramienta han sido la dificultad para encontrar lexicones completos y válidos para realizar un análisis de sentimiento en catalán y en castellano, y el haber sido utilizada anteriormente en otros estudios analizando ámbitos de elecciones similares a las de este documento (Tumasjan A. et al, 2010; Yu et al., 2008). Tras analizar los tweets de cada partido se realizó la media de sus valores en cada una de las dimensiones emocionales analizadas por la herramienta (en concreto se usó la que mide el sentimiento positivo), obteniendo así los valores de sentimiento por cada partido. Una vez reunida toda esta información se dio paso al análisis de correlación con el porcentaje de votos.

En cuanto a la correlación, el escenario de las elecciones de Cataluña ofrece unos resultados atípicos en comparación con estudios similares, como se puede ver en la Figura 11 y en la Tabla 6, y en particular frente a los anteriores estudios del profesor Moro. En varias investigaciones (Moro et al., 2014) el número de menciones a los partidos ha servido como indicador más o menos estable del porcentaje de votos recibido, mientras que en el presente caso la correlación llega a ser negativa. Viendo la Figura 11 se observa cómo Junts pel Sí (JxSi), que es el partido que más votos ha aglutinado, presenta el mínimo de menciones, al contrario que el Partido Popular (PP), que presenta el mayor número de menciones y menor número de votos. Una vez más el escenario de confusión causado por la declaración de la independencia de Cataluña en la vida real se traslada a Twitter. Otra posibilidad que haya causado esta descorrelación puede ser la corta vida que todavía tiene la plataforma de Junts pel Sí en la red social de Twitter frente a otros partidos. Esto se traduce en poca notoriedad en las redes sociales, pocos seguidores y por tanto pocas menciones. Sin embargo, el marco de las elecciones generales sigue el patrón de los anteriores estudios, llegando a obtener una correlación por encima del 90% en el caso de las menciones. Podría pensarse que la actividad en redes tiene lugar por medio de los constituyentes de la coalición Junts pel Sí y menos como coalición en sí. Sin embargo, observando los datos extraídos de Twitter se comprobó cómo tanto Convergents – CDC (la cuenta de Twitter de Convergència Democràtica de Catalunya, participante en la colación) y Esquerra Republicana (la cuenta de Esquerra Republicana) tienen un número considerablemente inferior comparados con la cuenta Junts pel Sí.

Asimismo se continúan observando, al igual que en anteriores estudios, los fenómenos de sobrerrepresentación de ciertos partidos políticos (CUP en las elecciones catalanas y PODEMOS en las generales tienen las comunidades más grandes de la muestra y no reúnen el mayor número de votos) e infrarrepresentación de otros (el PP serviría de ejemplo). Las comunidades obtenidas mediante la relación Hashtag han dado buenos resultados en los dos escenarios de campaña (alrededor del 70% de correlación, ver Tabla 6), manteniéndose estables en sus valores. Es necesario remarcar que el tamaño de la comunidad de Hashtags se calcula únicamente a través del número de usuarios de cada

partido, es decir, desechando los hashtags, ya que los nodos del grafo que se genera mediante esta relación son tanto usuarios como hashtags. Si no se hiciera esta criba el análisis sería incongruente frente a las otras comunidades (Retweet y Quote), ya que éstas se hacen únicamente contabilizando usuarios. Por otro lado, la ponderación del número de menciones a través del número de seguidores y el sentimiento positivo también ha supuesto una mejora de la correlación con el número de votos.

No obstante, los resultados siguen teniendo un error absoluto medio (MAE, es una métrica para medir la precisión de las predicciones), ver Tabla 6, demasiado elevado como para poder utilizarlos de herramienta predictiva. Además, en líneas generales estas aproximaciones no pueden competir con otros modelos obtenidos de otras fuentes como pueden ser las encuestas del CIS (Centro de Investigaciones Sociológicas). Sin embargo, la combinación entre número de menciones, seguidores y sentimiento positivo obtuvo en las elecciones generales un 94% de correlación frente al 91% de la encuesta del CIS en Octubre de 2015 (Figura 14), pero esto no puede considerarse una regla general, puesto que habría que realizar posteriores análisis observando si existe una pauta marcada.

A continuación se muestran las gráficas obtenidas del cálculo de la correlación. En cada gráfica cada punto representa un partido/comunidad de partidarios en Twitter, el eje Y representa el porcentaje de votos obtenidos en las elecciones y el eje X representa los tamaños de las comunidades de partidarios o el número de menciones a cada partido (se indica cada caso en el título de los ejes). La línea azul representa la línea de tendencia de cada gráfica, pudiendo observar si la tendencia del conjunto de datos crece o decrece (en nuestro caso si la correlación es positiva o negativa). Los aspectos que más caben destacar de las gráficas, son, en la Figura 10, la gran separación entre Junts Pel Sí (por encima de la línea de tendencia) y los demás partidos (por debajo de la línea de tendencia). Este fenómeno se suaviza levemente en el marco de las elecciones generales (ver Figura 12 y Figura 13) hasta casi alinearse con la línea de tendencia salvo el caso del PSOE.

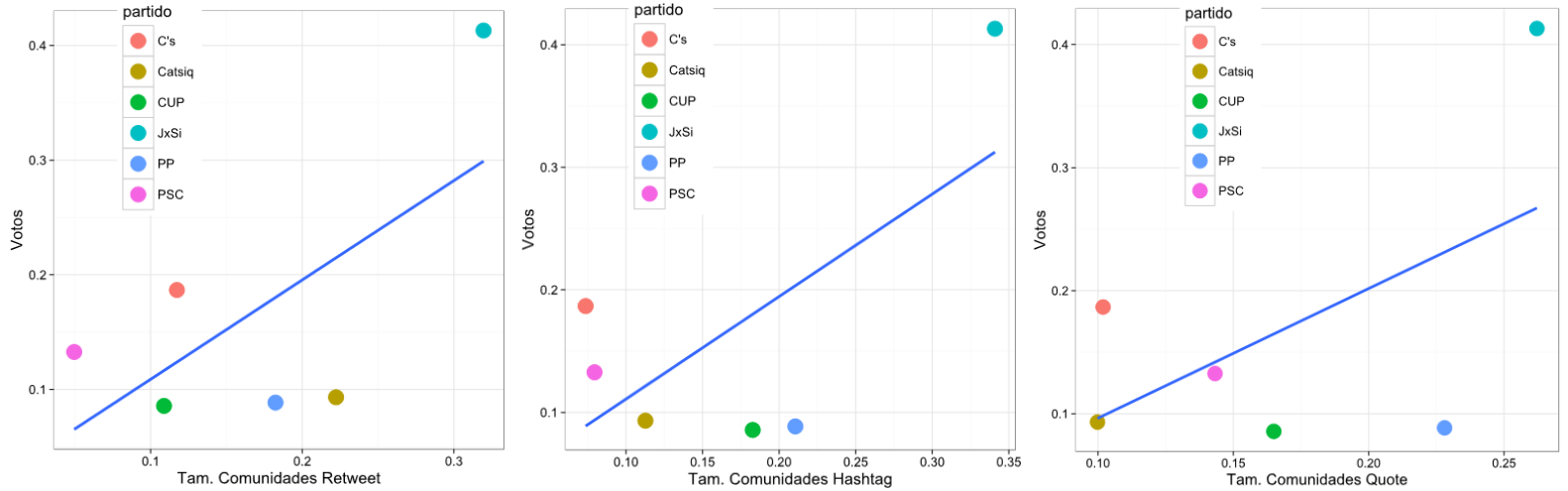

Figura 10: Correlación entre el tamaño de cada comunidad y el porcentaje de votos obtenido en las elecciones de Cataluña de 2015. Se han estudiado la relación de Retweet (izquierda), la relación de Hashtag (centro) y la relación de Quote.

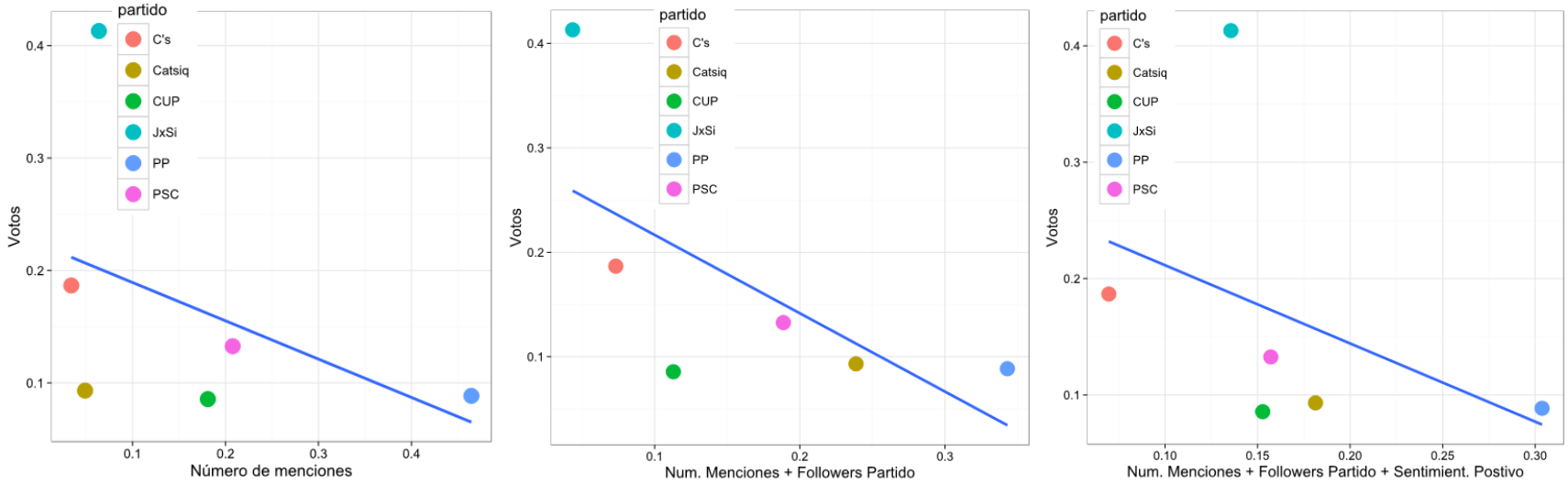

Figura 11: Correlación entre el número de menciones (izquierda) el porcentaje de votos obtenido en las elecciones de Cataluña de 2015. Se han estudiado también formas ponderadas por el número de followers (centro) y el número de followers y el sentimiento positivo de los tweets de cada partido (derecha).

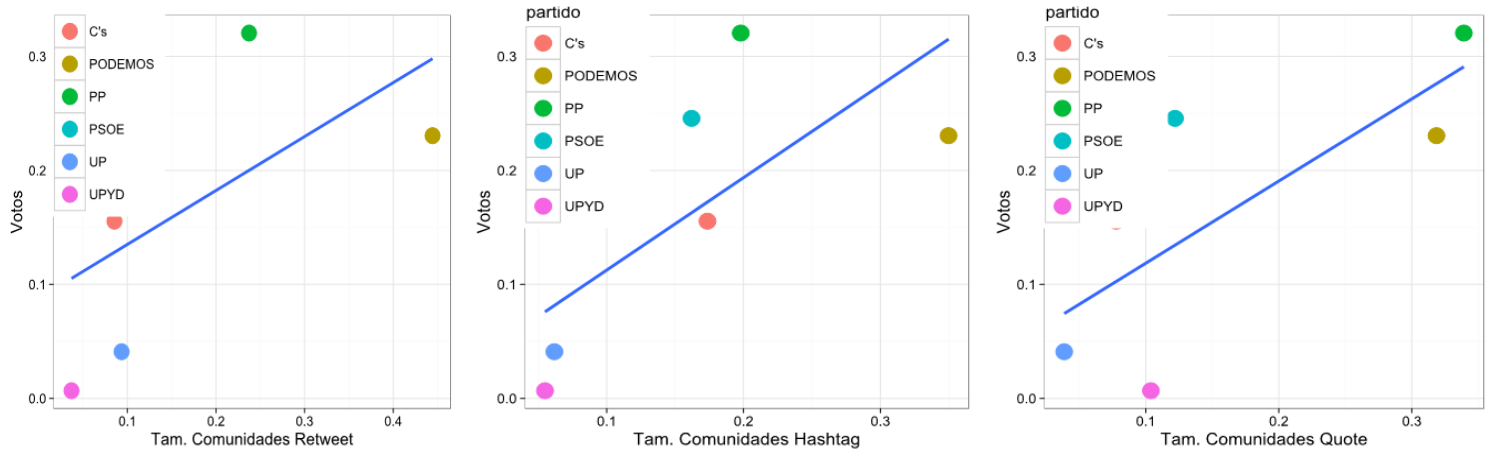

Figura 12: Correlación entre el tamaño de cada comunidad y el porcentaje de votos obtenido en las elecciones de generales de 2015. Se han estudiado la relación de Retweet (izquierda), la relación de Hashtag (centro) y la relación de Quote.

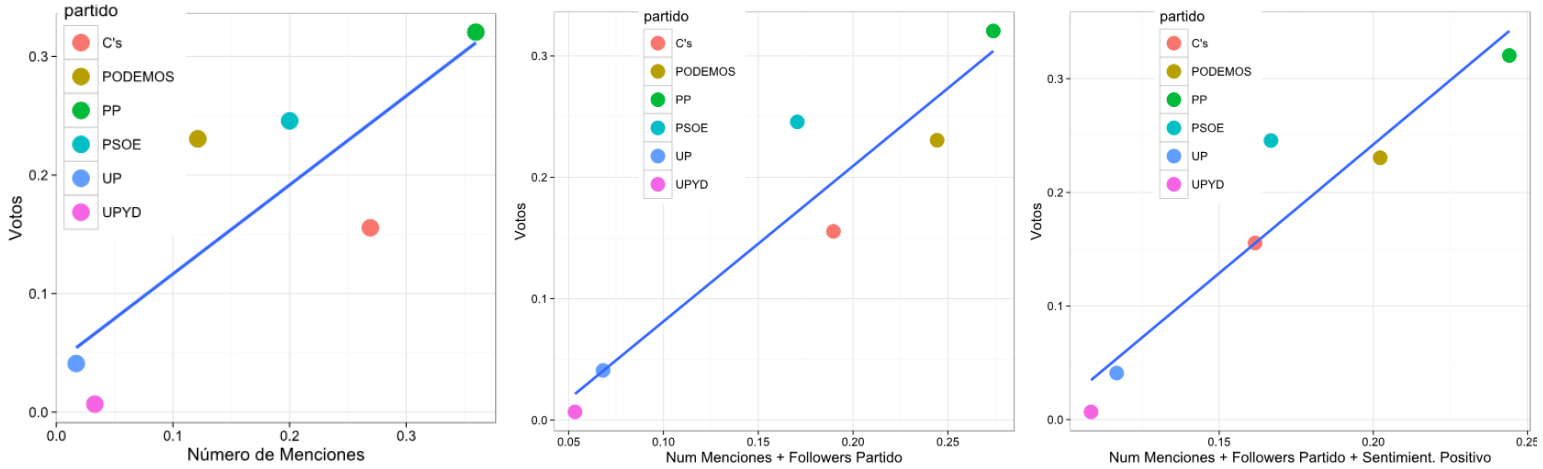

Figura 13: Correlación entre el número de menciones (izquierda) el porcentaje de votos obtenido en las elecciones generales de 2015. Se han estudiado también formas ponderadas por el número de followers (centro) y el número de followers y el sentimiento positivo de los tweets de cada partido (derecha).

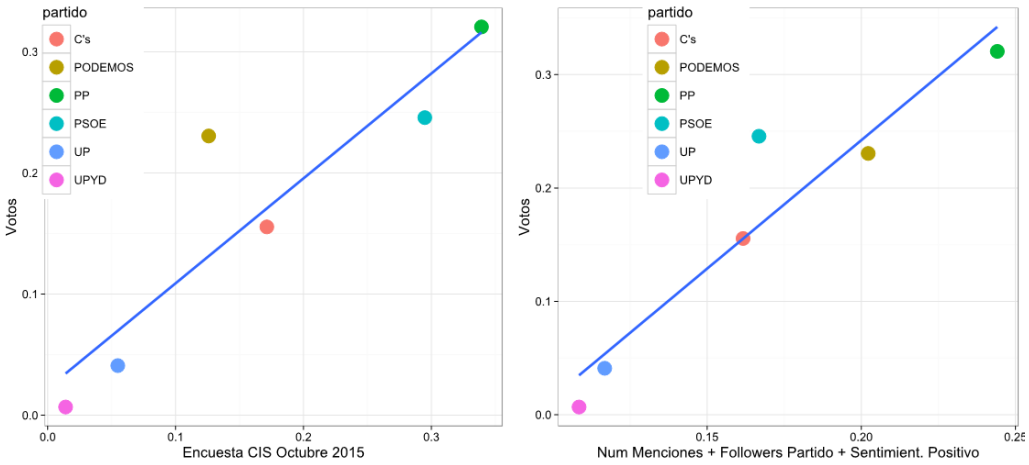

Figura 14: Comparativa entre la encuesta del CIS de Octubre de 2015 (izquierda) con un 91% de correlación (Pearson) y un 3% de MAE y entre el número de menciones ponderado (derecha) con un 94% de correlación y un 6% de MAE.

Elecciones de Cataluña				
Medida	Correlación Pearson	Interv. Confianza Pearson	Correlación Spearman	MAE
Comunidad Retweet	0.65	[-0.17, 0.93]	0.37	0.08
Comunidad Quote	0.55	[-0.31, 0.91]	0.08	0.07
Comunidad Hashtag	0.67	[-0.13, 0.94]	-0.08	0.07
Número de Menciones	-0.43	[-0.88, 0.45]	-0.48	0.18
Menciones + Followers partido	-0.66	[-0.94, 0.15]	-0.65	0.1
Menciones + Sentimiento + Followers partido	-0.40	[-0.87, 0.48]	-0.6	0.13
Elecciones Generales				
Medida	Correlación Pearson	Interv. Confianza Pearson	Correlación Spearman	MAE
Comunidad Retweet	0.58	[-0.27, 0.92]	0.77	0.09
Comunidad Quote	0.75	[0.02, 0.95]	0.77	0.06
Comunidad Hashtag	0.71	[-0.06, 0.95]	0.71	0.06
Número de Menciones	0.82	[0.20, 0.97]	0.77	0.05
Menciones + Followers partido	0.93	[0.60, 0.98]	0.82	0.04
Menciones + Sentimiento + Followers partido	0.94	[0.65, 0.99]	0.94	0.06

Tabla 6: Correlación entre las distintas medidas de intención de voto y el resultado oficial. Para cada medida se muestra también el intervalo de confianza de la correlación y el MAE.

4.5 Intención de voto y geolocalización

Otro enfoque interesante radica en el análisis geolocalizado de la intención de voto. Una pequeña parte de nuestra descarga de tweets durante las elecciones generales (0,001% sobre el total de tweets) eran geolocalizados, a los cuales se les ha extraído el usuario y se ha contrastado sobre el resultado de la detección de comunidades de partidarios. Una vez identificados se les ha marcado según la comunidad de partidarios a la que pertenecen y la comunidad autónoma en la que han realizado los mensajes. Para la visualización se ha utilizado la aplicación CartoDB, una herramienta puntera par la creación de mapas, visualizaciones y análisis de datos. El resultado se puede ver en la Figura 15. La predicción de votos se ha realizado sobre los puntos más calientes del mapa de calor. La selección se realizó localizando los tweets por comunidad autónoma a través de la herramienta de información geográfica QGis, escogiendo las comunidades con mayor densidad de tweets geolocalizados. Los puntos analizados son Madrid (22% de los tweets geolocalizados), Andalucía (19%), Cataluña (11%) y finalmente Valencia (10%)

Los resultados han sido dispares (ver Figura 16): Valencia y Andalucía muestran una correlación alrededor del 70%, Cataluña del 80% y Madrid del 30%. El escaso número de tweets geolocalizados sobre la muestra puede haber provocado un aumento de los sesgos, sin embargo puede ser interesante en futuras investigaciones de esta índole centrarse en el aspecto geográfico durante campañas electorales.

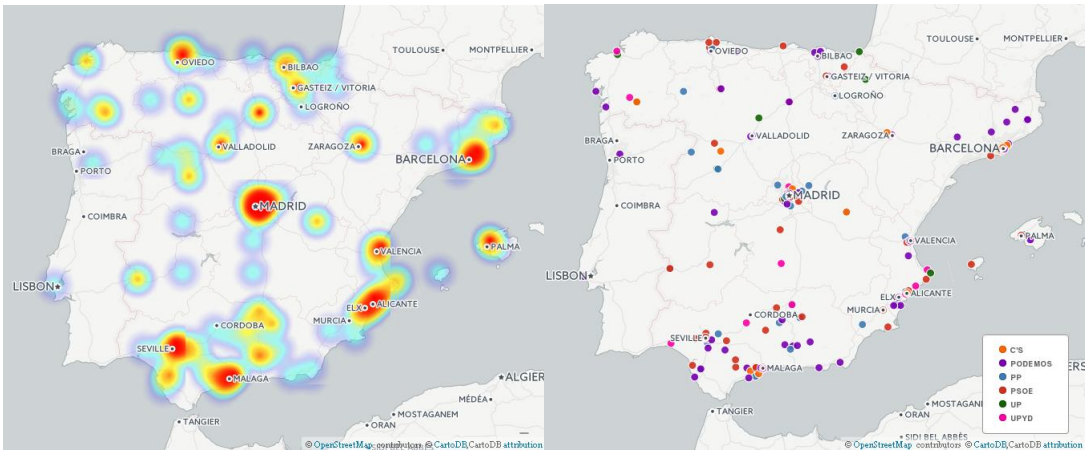

Figura 15: Mapa de calor (izquierda) con los puntos en los que se concentran los tweets geolocalizados y mapa con los tweets clasificados por partido (derecha).

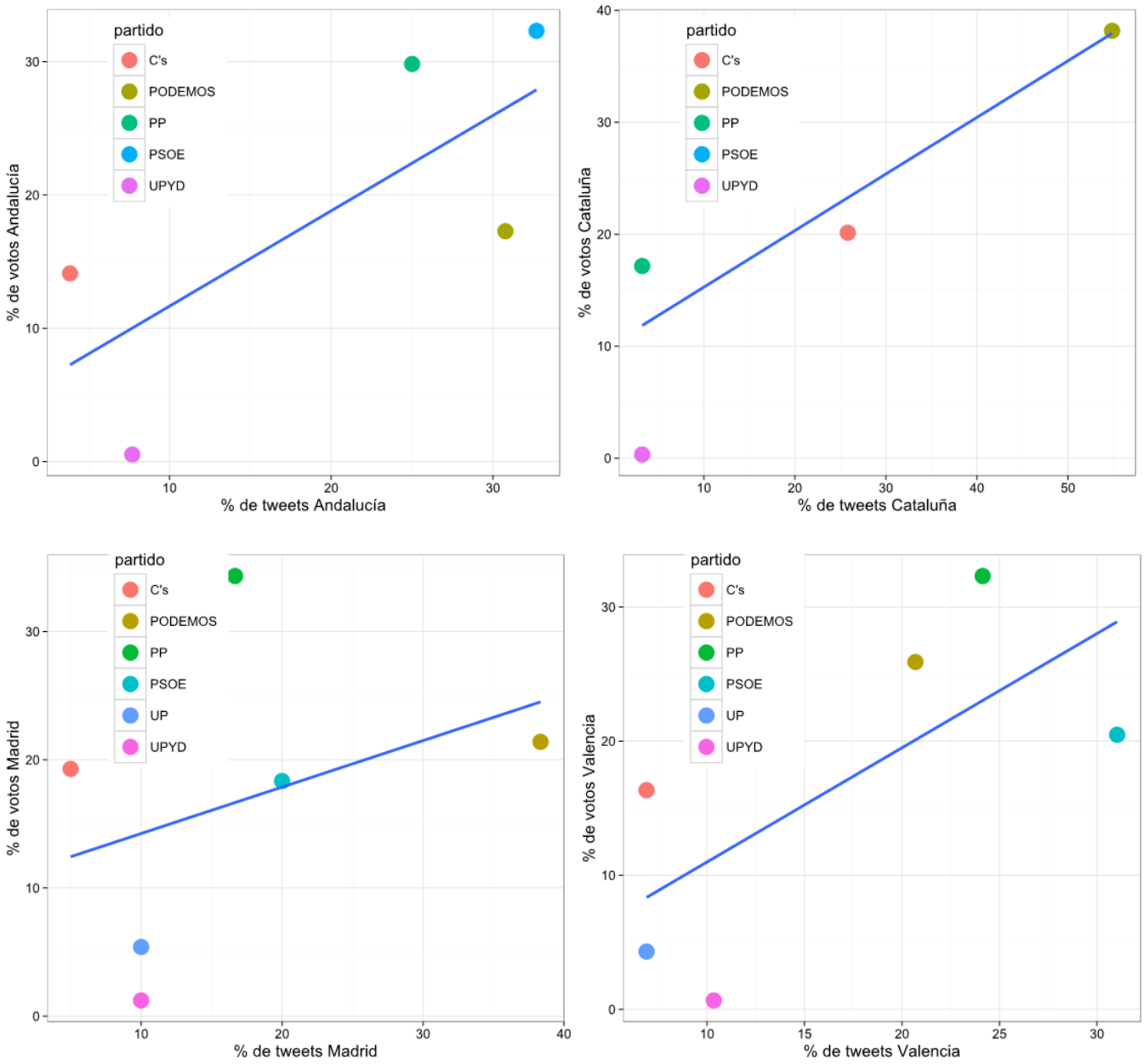

Figura 16: Correlación entre el número de tweets geolocalizados por partido y los votos por comunidad autónoma.

4.6 Análisis de la polaridad de los tweets a través de clasificadores dentro del marco de la independencia de Cataluña

Tal como hemos planteado en la introducción, hemos llevado a cabo un análisis de polaridad, para ello hemos escogido los tweets referidos a la independencia de Cataluña. El análisis se hará mediante un clasificador no supervisado basado en los modelo de Teoría de Respuesta al Ítem (IRT). Esta teoría ha sido desarrollada recientemente en el campo de la psicometría, utilizando modelos matemáticos para describir la relación entre el nivel de habilidad de un sujeto (por ejemplo el cociente intelectual, CI) y la probabilidad que éste dé una respuesta correcta a un ítem de un test psicológico. En el marco político el modelo funciona utilizado una matriz en la que hay v (*voters*) filas representando a los votantes y b (*bills*) columnas representando cada una de las votaciones. Cada celda de la matriz representa la decisión de un legislador en una determinada votación, que puede tener un valor de 1 ó 0 (sí o no). Además, para aplicar el modelo se necesitan hacer dos supuestos: que todos o la mayoría de los votantes pueden describirse posicionándolos a lo largo de una única línea continua y que la posición de cada votante dentro de la línea influye en sus votos. El modelo se representa como una regresión logística:

$$P(y^{vb} = 1) = logit(b1^{b} * x^{v} - b0^{b})$$

Figura 17: Regresión logística utilizada para el modelo.

La probabilidad $P(y^{vb} = 1)$ es la probabilidad de que un legislador en una determinada votación vote a favor, x es la posición ponderada de cada votante v, b0 es la dificultad de votar si por cada votación b, y b1 es el grado con el que la posición de cada votante afecta para votar a favor de cada votación b. De esta forma, b1 tomará valores negativos o positivos respectivamente en los casos que votantes de una u otra ideología voten o no a favor en una determinada votación (Heimann & Danneman, 2014; Ravichandran et al., 2015).

Para el caso de este documento, los legisladores serían usuarios de Twitter, y las votaciones bigramas de palabras utilizadas en los tweets. De esta forma, el uso o no de ciertos bigramas de palabras será indicativo del posicionamiento ideológico de cada usuario analizado. Para esto se han obtenido dos conjuntos de datos a través de hashtags: los tweets de carácter independentista (#independencia, #governemnos, etc) etiquetados con un 1 y los tweets de carácter antiindependentista (#cataluñaespespaña, #catalanyespañol, etc) etiquetados con un -1. Cada grupo contiene 6325 tweets. Los bigramas se formarían a través de la matriz de términos formada por las palabras utilizadas en los tweets, tras haber realizado un filtro de stop words y palabras que no aportan nada significativo.

La Figura 18 muestra los bigramas utilizados. Cada punto en el gráfico es un bigrama, el eje X muestra su grado de dificultad con el que aparece en la muestra de tweets y el eje Y representa el nivel de discriminación, es decir, el nivel de polarización. Estas dos medidas tienen un alto grado de correlación, por lo que los bigramas que tengan un alto valor

discriminativo y un alto grado de dificultad serán los más significativos para clasificar a los usuarios. La forma en flecha del gráfico muestra esa significancia, siendo los bigramas comunes que no son discriminatorios los que se agrupan en la "punta de la flecha". En el gráfico se han etiquetado los bigramas más polarizados, en la parte positiva del eje Y están los independentistas, mientras que en la parte negativa están los antiindependentistas.

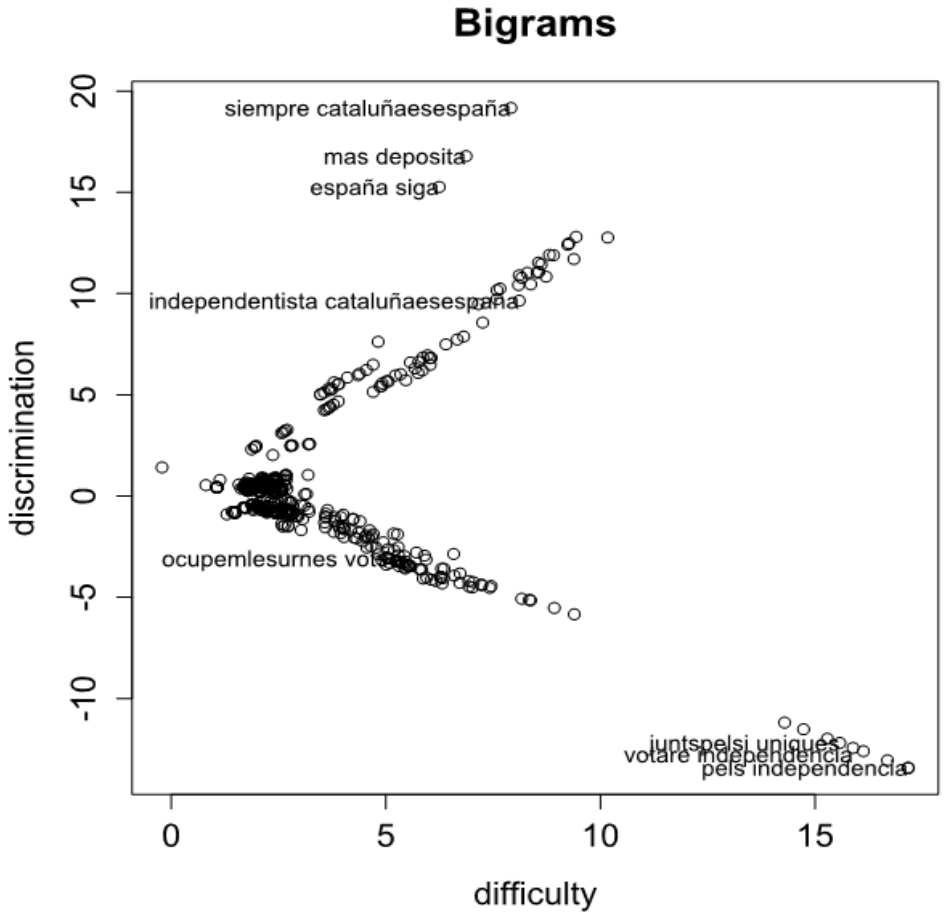

Figura 18: Bigramas utilizados en el modelo IRT según su nivel de discriminación (positivo para una clase y negativo para otra) y su grado de dificultad de aparición.

En cuanto a la clasificación de usuarios, se etiquetaron según su prodencia, es decir, según se encontraran en el conjunto de tweets obtenidos mediante hashtags antiindependentistas o independentistas. El método es el siguiente: como los tweets se etiquetaron previamente con un valor 1 ó -1 según el conjunto, para cada usuario se hace la media de los valores etiquetados de sus tweets, por lo que si la media resulta mayor que 0 se le etiqueta como independentista, menor que 0 antiindependentista y si es igual a 0 ese usuario se considera indeterminado y se descarta. Esta acción está orientada en base a una futura evaluación de la tasa de acierto del clasificador. La Figura 19 muestra el resultado de aplicar el modelo IRT frente al etiquetado manual, los usuarios que aparecen en los bordes del gráfico son lo que tienen mayor índice de polaridad. Estos usuarios son, en el lado

independentista, “Marisa #JuntsPelSí”, “Associació DNI.cat” y “Laksmí‖★‖”, y en el lado antiindepentista “PPMunicipioElEspinar”, “PP Català” y “Partido Popular”. También se puede observar que existen ciertos usuarios clasificados incorrectamente, teniendo este clasificador una tasa de acierto del 91%, frente a otros clasificadores no supervisados como K-means (80%) o mclust (75%) (Fraley C. et al, 2012).

Como se puede observar en la Figura 19, los usuarios se han posicionado a lo largo de una línea recta en la que su posición en la misma supone un posicionamiento ideológico, tal y como se dijo anteriormente al hablar de la aplicación del ITR en las ciencias políticas. Se ha representado como una función ascendente simplemente con el fin de ayudar a la visualización.

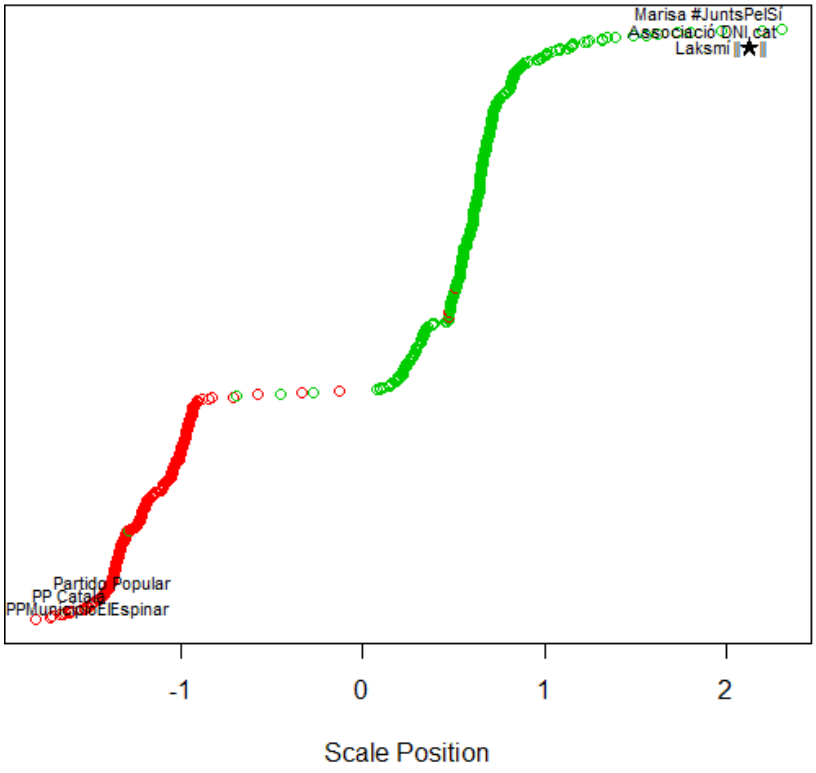

Figura 19: Usuarios cuyos tweets han sido analizados mediante IRT, el grupo del conjunto con hashtags antiindependentistas (rojo) e independentistas (verde). El eje x muestra el nivel de polaridad (positiva o negativa) de cada uno de los usuarios.

5 Conclusiones y trabajo futuro

5.1 Conclusiones

Las principales conclusiones obtenidas a partir de los resultados del presente estudio han sido las siguientes:

- **Relevancia de las nuevas relaciones de organización**. Anteriores estudios tomaron únicamente la relación de Retweet como relación significante a la hora de formar grafos. En nuestro estudio se ha propuesto también la utilización de las relaciones de Hashtag y Quote como relevantes. Los grafos generados mediante estas relaciones muestran también una polarización política evidente (grupos de partidarios de cada partido político) tal y como se ha observado en anteriores investigaciones basadas en el análisis de interacciones en Twitter.
- **Calidad de las predicciones**. El grado de correlación entre el tamaño de las comunidades y los votos ha sido ligeramente superior en la mayoría de los casos con las relaciones propuestas de Hashtag y Quote, pero sigue siendo demasiado bajo para considerarse relevante como indicador. Respecto al grado de correlación a través del número de menciones, la combinación con el número de followers de cada partido junto con el sentimiento positivo medido de los tweets ha supuesto una leve mejora de la correlación (en el caso de las elecciones generales incluso respecto al CIS). Sin embargo, todavía no se puede considerar este resultado como comportamiento universal, ya que queda contrastarlo con otros análisis distintos centrados en campañas electorales.
- **Discrepancia respecto a resultados anteriores**. Las elecciones de Cataluña, con la problemática implícita de la declaración de independencia, han podido desencadenar una ruptura de las tendencias generales de anteriores investigaciones. En concreto, el grado de correlación obtenido entre el número de menciones de cada partido frente al número de votos respectivo ha resultado inverso, a pesar de que este método ha dado buenos resultados anteriormente y ha sido utilizado recurrentemente como indicador de intención de voto.
- **Comunicación entre comunidades**. El análisis de Moro (Moro et al., 2014) revelaba un flujo de las comunicaciones en las comunidades tanto interno como externo, siendo este último más fuerte entre comunidades que comparten cierta afinidad política. Nuestro estudio muestra resultados que se desvían de esta tendencia general en determinados casos, por lo se podría sugerir el interés de añadir una dimensión

más, el antagonismo político y/o el surgimiento de un partido al que apuntan todos los focos mediáticos (PODEMOS).

- **Polaridad de los discursos.** A través de la clasificación obtenida mediante el sistema IRT se puede ver como en el marco de las elecciones catalanas se conforman dos sentimientos contrarios: a favor y en contra de la independencia. Lo ventajoso de este método es que permite discriminar tanto usuarios como bigramas o términos utilizados en tweets por esos mismos usuarios, además de proporcionar mejores resultados comparado con otros algoritmos de clasificación de textos no supervisados.

5.2 Trabajo futuro

Hay ciertos aspectos a tener en cuenta a la hora de intentar ampliar o mejorar los resultados obtenidos en este documento en futuras investigaciones:

- **Eliminar ruido de los datos extraídos.** En Twitter existen multitud de cuentas falsas, spammers o bots. La propia empresa considera que menos del 5% de sus usuarios activos mensajes son fakes (falsos). Muy probablemente, en los datos utilizados en este documento se hayan filtrado este tipo de cuentas, por lo que en futuras investigaciones habrá que tomar medidas de detección de bots.
- **Aumentar la duración de la extracción de datos.** Al utilizar recursos limitados para el desarrollo de este estudio el seguimiento de las campañas electorales sobre Twitter ha durado alrededor de una semana. La duración óptima de la extracción de datos sería el periodo durante el que transcurre la campaña electoral, estipulado por la ley como los quince días antes de las elecciones. Utilizando un gran conjunto de datos en el futuro, se podrían evitar ciertos sesgos producidos por tendencias temporales fruto de la campaña. Además, en el caso de los tweets geolocalizados, tener una muestra mayor de los mismos permitiría mejorar el análisis de correlación con los votos obtenidos en cada una de las Comunidades Autónomas de España.
- **Mejorar el sistema de recolección de datos.** Otra posible mejora de cara al futuro en el caso que se extraigan una gran cantidad de datos, es el uso de tecnologías basadas en el Big Data, como son las bases de datos no estructuradas (NoSQL), pensadas para escribir miles de registros por segundo y en las que no se suelen modificar los datos (perfectas para realizar streaming sobre Twitter). Otra mejora sería el uso del sistema Hadoop, basado en el paradigma map-reduce, que permite el procesamiento paralelo y masivo de gran cantidad de datos a un bajo coste.

Referencias

Barabási, A.-L. (2003). Linked: How Everything is Connected to Everything Else and what it Means for Business, Science, and Everyday Life. Plume.

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. Science, 286(5439), 509-512. http://doi.org/10.1126/science.286.5439.509

Blondel et al. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. http://doi.org/10.1088/1742-5468/2008/10/P10008

Cohen et al. (2000). Resilience of the Internet to random breakdowns. Physical Review Letters, 85(21), 4626-4628. http://doi.org/10.1103/PhysRevLett.85.4626

Congosto et al. (2013). Twitter y política: Información, opinión y ¿Predicción? Recuperado a partir de http://markov.uc3m.es/~emoro/ps/evoca.pdf

Conover et al. (2012). Partisan Asymmetries in Online Political Activity. EPJ Data Science, 1(1). http://doi.org/10.1140/epjds6

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection (p. 104(1):36–41). Proceedings of the National Academy of Sciences of the United States of America.

Fraley C. et al. (2012). MCLUST Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation (Technical Report No. 597). Department of Statistics, University of Washington.

Gayo-Avello, D. (2012). «I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper» -- A Balanced Survey on Election Prediction using Twitter Data. arXiv:1204.6441 [physics]. Recuperado a partir de http://arxiv.org/abs/1204.6441

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821-7826. http://doi.org/10.1073/pnas.122653799

Heimann, R., & Danneman, N. (2014). Social Media Mining with R. Packt Publishing.

Jackman S. et al. (2008). Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University, Stanford, California. Recuperado a partir de http ://CRAN.R-project . org/ package=pscl

Jungherr et al. (2011). Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. «Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment». Social Science Computer Review, 894439311404119. http://doi.org/10.1177/0894439311404119

K. Knibbs. (2015). The FBI Says Retweets Are Endorsements. Recuperado a partir de http://gizmodo.com/the-fbi-says-retweets-are-endorsements-1731526051

Kirkland, S. (2014). Retweets are endorsements at NPR and AP, but not at NYT. Recuperado a partir de http://www.poynter.org/2014/retweets-are-endorsements-at-npr-and-ap-but-not-at-nyt/258240/

Lorente, U. (2015). El ejército de Podemos en Twitter: más de 20 «soldados» para gestionar la cuenta oficial del partido. Recuperado a partir de http://vozpopuli.com/actualidad/58270-el-ejercito-de-podemos-en-twitter-mas-de-20-soldados-para-gestionar-la-cuenta-oficial-del-partido

Moro et al. (2014). La democracia del siglo XXI. Política, medios de comunicación, internet y redes sociales. Actas de las II Jornadas españolas de ciberpolítica, 28 de mayo de 2013. Madrid: Centro de Estudios Políticos y Constitucionales.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E, 69(2), 26113. http://doi.org/10.1103/PhysRevE.69.026113

Patrick A. et al. (2015). igraph: Network Analysis and Visualization. Recuperado a partir de https://cran.r-project.org/web/packages/igraph/index.html

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks (long version). arXiv:physics/0512106. Recuperado a partir de http://arxiv.org/abs/physics/0512106

Ravichandran et al. (2015). Intelligent Topical Sentiment Analysis for the Classification of E-Learners and Their Topics of Interest. The Scientific World Journal, The Scientific World Journal, 2015, 2015, e617358. http://doi.org/10.1155/2015/617358, 10.1155/2015/617358

Tumasjan A. et al. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.

Twitter España. (2015). ¡Estrenamos emojis para las elecciones del #20D! [microblog]. Recuperado a partir de https://twitter.com/TwitterSpain/status/669798422648635392

Yu et al. (2008). Exploring the Characteristics of Opinion Expressions for Political Opinion Classification. The School of Information Studies Faculty Scholarship. Recuperado a partir de http://surface.syr.edu/istpub/34

Zarrella, D. (2010). Can Twitter predict elections? Recuperado a partir de http://danzarrella.com/new-data-can-twitter-predict-elections.html

Anexos

A Estadísticas de Twitter sobre el conjunto de datos de las Elecciones de Cataluña

Usuarios	Número de menciones
El Español	797
Partido Popular	498
EL PAÍS	390
EL MUNDO	353
ABC.es	267
eldiario.es	205
324.cat	191
VOX	180
Los Mejores Vines	161
CUP #Governemnos	151

Hashtags	Número de tweets
#27s	248794
#eleccionescatalanas	140159
#juntspelsi	15735
#independencia	11226
#27s2015	9992
#governemnos	9204
#votapermi	8613
#guanyemjunts	7398
#catalunya	7116
#27stv3	6949

Tweets	Número de retweets
Todo preparado en el colegio electoral para que vaya a votar Miquel Iceta #27S http://t.co/XVR0mI8BLf	2518
Un chaval con una bandera española expulsado del colegio electoral durante el voto de Mas ##27s https://t.co/bjuUAr2Ep8	2361
Los independentistas en las urnas.... #eleccionescatalanas http://t.co/mQjGUxMNR5	1348
I que només els petons ens tanquin la boca Guanyarem! #27S #independencia #pobresazero #proucorrupció #dempeus #CUP http://t.co/Tq0S6T7J7d	1289
A ver si lo entiendo @CiudadanosCs que criticaba la celebración de tres #eleccionescatalanas pide otras hoy? #LoFlipo http://t.co/MLcem2zjtX	1262
Els catalans residents als Estats Units, Xile, Mèxic i Xangai no han pogut votar http://t.co/N4Wb8X4h8I #27S http://t.co/AO36jVQET3	1142
Este es el joven que ha sido expulsado durante el voto de Mas al sacar una bandera española #eleccionescatalanas http://t.co/Nraj68ahZC	1126
Verás cuando Córdoba haga el referéndum #eleccionescatalanas http://t.co/vDtZC1bdMW	1099
Buenos días! Voy a votar para que España siga unida. #CatalánYEspañol #27S http://t.co/lW5Dv3ms23	928
Gane quien gane las #eleccionescatalanas, Cataluña seguirá siendo parte de Alemania. Como el resto de Europa.	923

B Estadísticas de Twitter sobre el conjunto de datos de las Elecciones Generales

Usuarios	Número de menciones
PSOE	909
eldiario.es	779
Nathy	578
SAC DA VIDA	532
El Español	515
Hay que pararlos	495
EL PAÍS	438
Publico.es	404
UPYD	366
Mariano Rajoy Brey	347

Hashtags	Número de tweets
#20d	189504
#podemos	56402
#ciudadanos	46415
#psoe	35099
#partidopopular	33577
#jornadadereflexion	21806
#eleccionesgenerales2015	17269
#upyd	16108
#votapsoe	11825
#caraacaral6	11242

Tweets	Número de retweets
-Yo voy a votar a Ciudadanos porque quiero probar algo nuevo. +¿Te has disparado alguna vez en el estómago? - No. +Mira, algo nuevo. Prueba.	1016
"Votaré al PSOE para que no gane el PP." "Tomaré cianuro en lugar de tirarme por el balcón.."	945
ATENCIÓN: La Junta Electoral obliga a los consulados a abrir sábado y domingo para votar desde el exterior #20D . Difusion por favor	895
'La fuerza renace' este 20D y es un honor no estar del lado oscuro ;) #UnPaisConSusFrikis https://t.co/wak7taUKyM	894
El mejor discurso de la historia tras el "I have a dream" de Luther King es de la número 1 de Ciudadanos en Sevilla. https://t.co/cSdJl10Kvs	853
SOMOS NUMERO 1!!!!!! No nos lo podemos creer!! Gracias a todos los que estáis confiando en Ghost Town!! #https://t.co/xLuGg5yCYM	753
Podemos ser lo mejor, o también lo peor, con la misma facilidad.	747
Así quedaría el congreso si hubiera circunscripción única: PP 100 PSOE 77 Podemos 72 Cs 49 IU 12 ERC 8 DL 8 PNV 4 Bildu 3 Pacma 3 UPyD 3 ...	727
Necesito que todo el que piense votar a ciudadanos vea esto, rt y difusión por favor https://t.co/k9F1zWIaPO	722
Reacciona con Serenidad, hay cosas que no puedes cambiar, lo único que podemos hacer es aceptarlas. #SiempreFiel https://t.co/uQ194hqklJ	720

C Comunidades de partidarios detectadas en los grafos

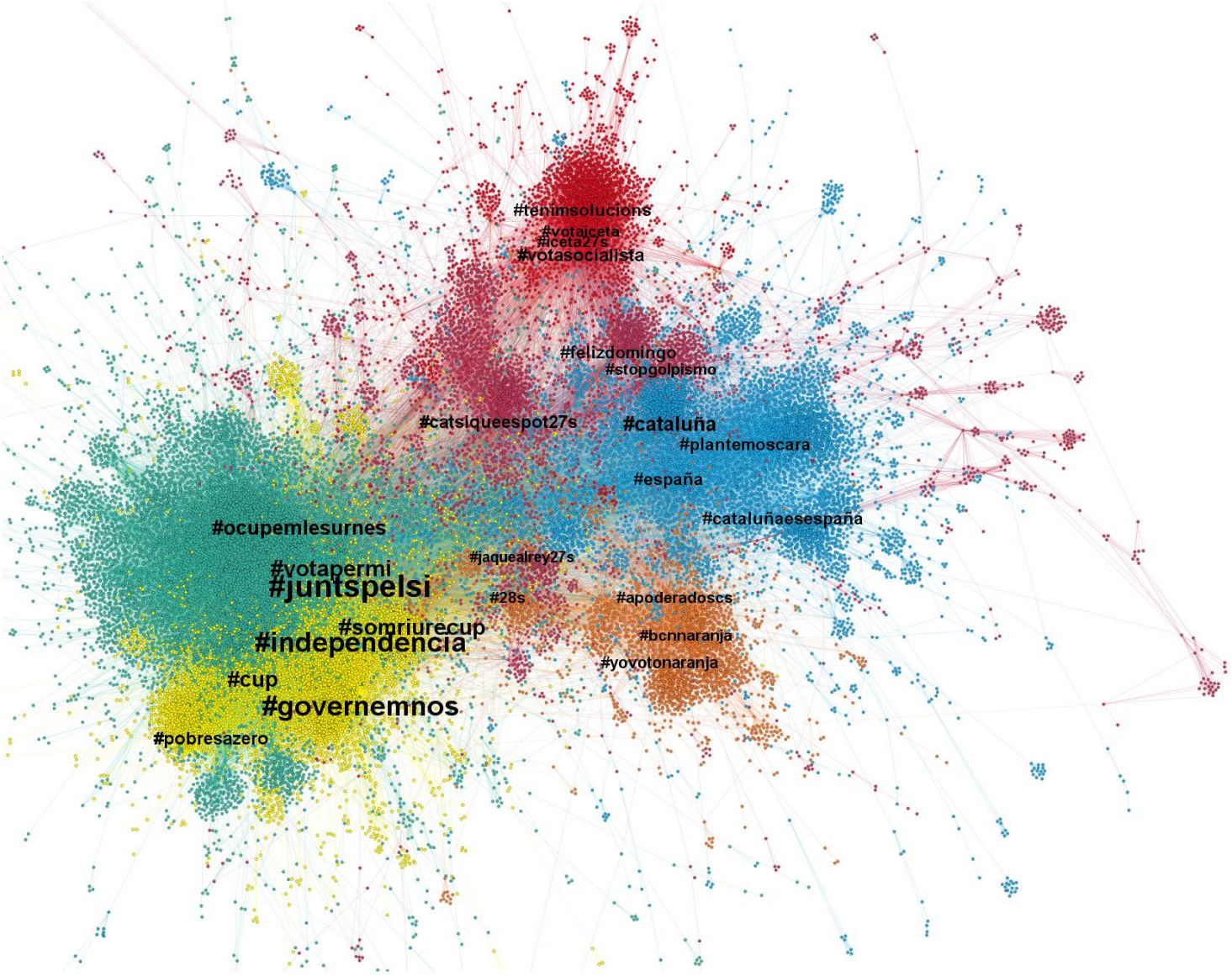

Figura 20: Grafo de Hashtags de las elecciones catalanas del 2015.

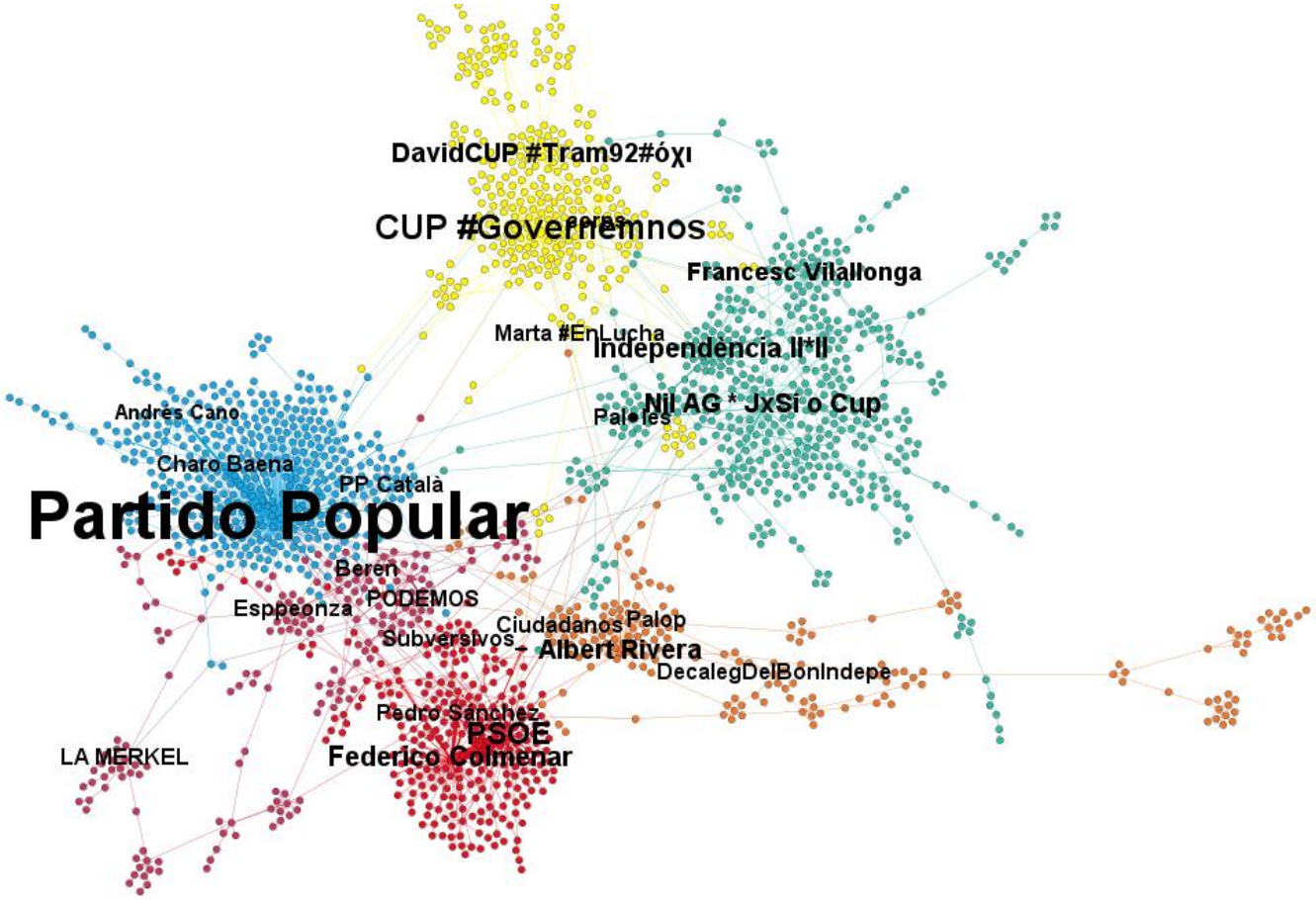

Figura 21: Grafo de Quotes (citas) de las elecciones catalanas del 2015.

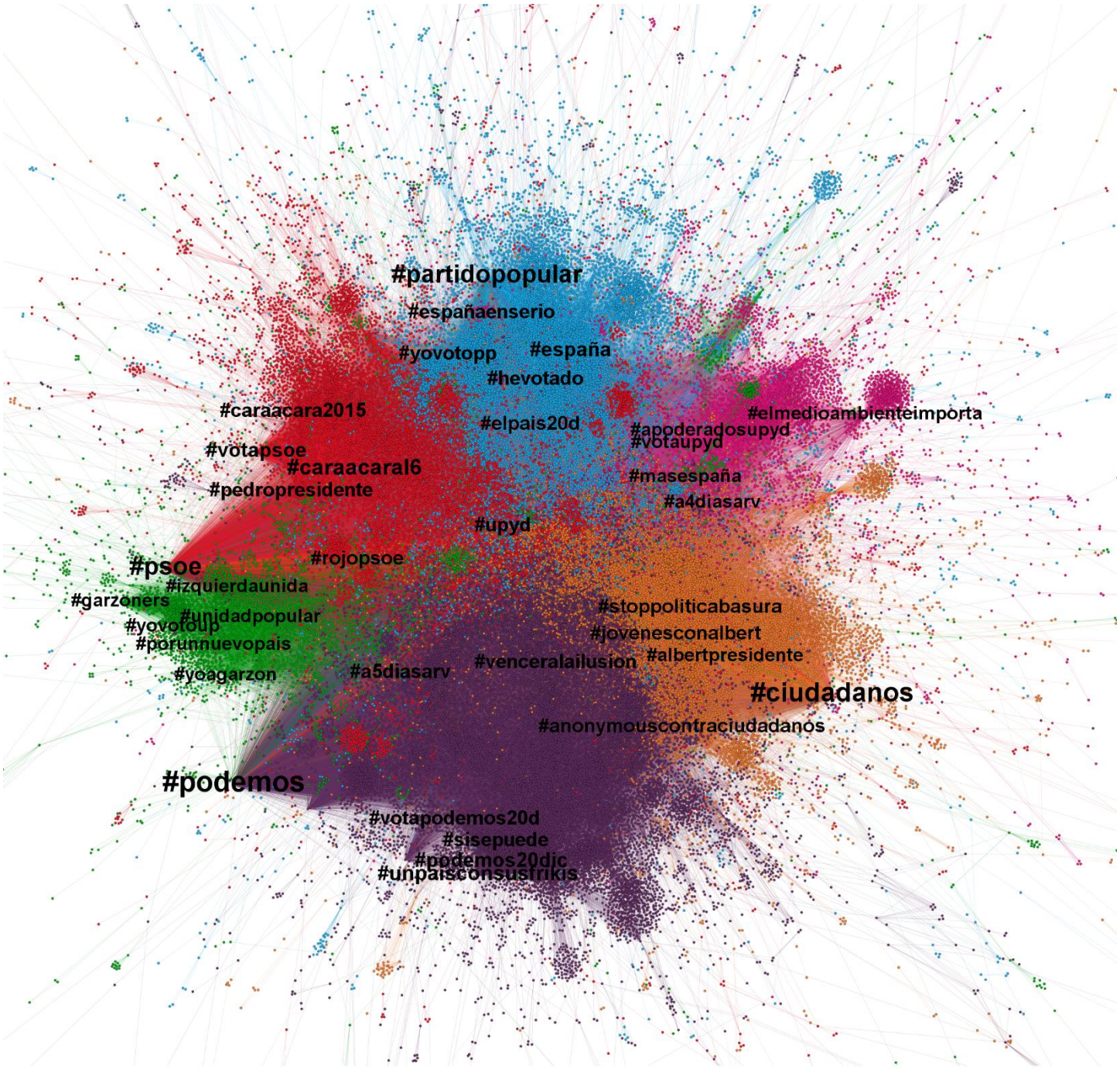

Figura 22: Grafo de Hashtags de las elecciones generales del 2015.

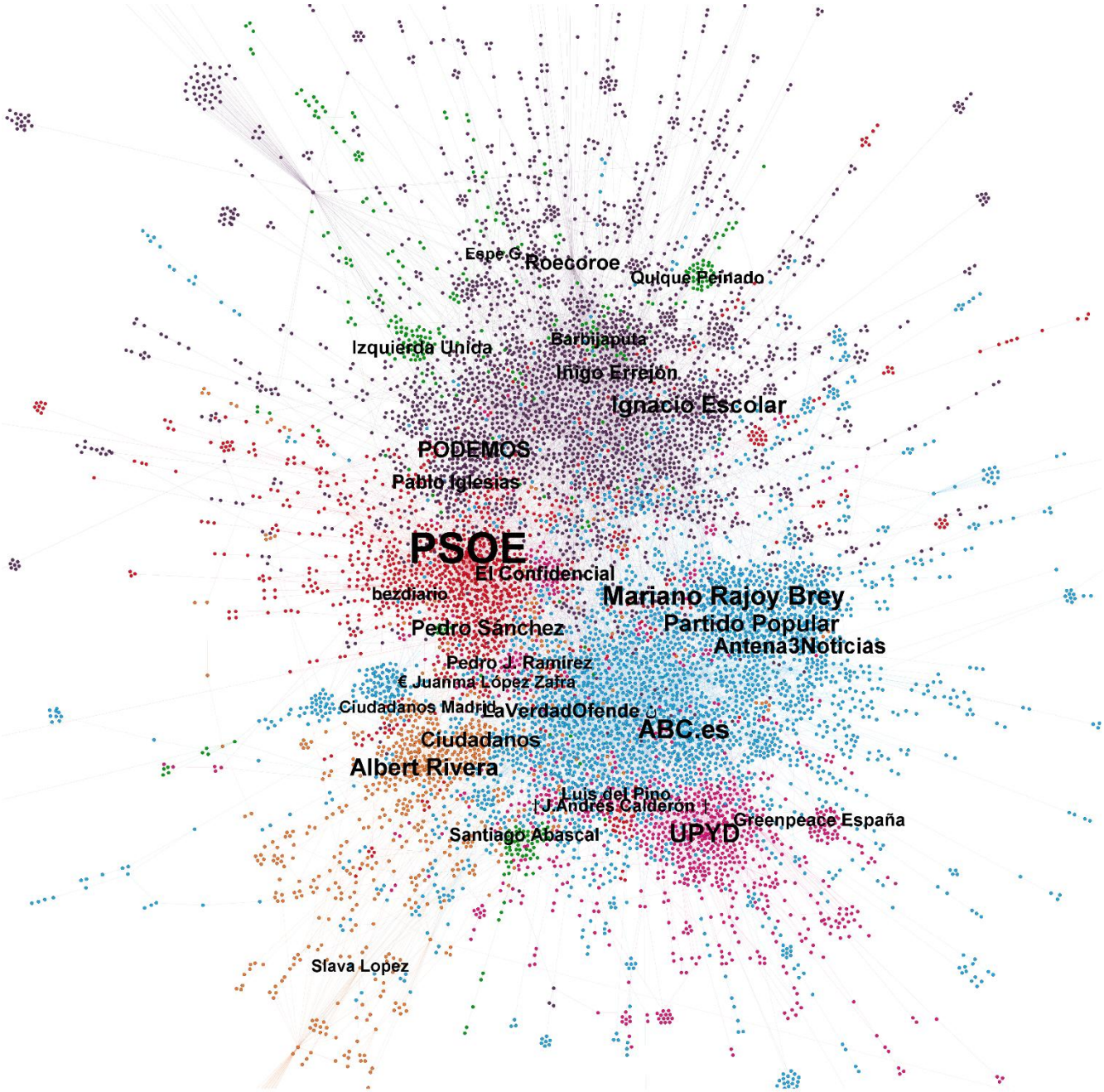

Figura 23: Grafo de Quotes (citas) de las elecciones generales del 2015.